

# PREDICCIÓN DEL NIVEL DE DIFICULTAD EN UNA PRUEBA ESTANDARIZADA DE COMPRENSIÓN DE LECTURA: APORTES DESDE LA PSICOMETRÍA Y LA PSICOLOGÍA COGNITIVA

*[Prediction of the difficulty level in a standardized reading comprehension test: contributions from cognitive psychology and psychometrics]*

by/por

[Article record](#)

[About authors](#)

[HTML format](#)

Brizuela, Armel ([armel9@gmail.com](mailto:armel9@gmail.com))

Montero-Rojas, Eiliana ([eilianamontero@gmail.com](mailto:eilianamontero@gmail.com))

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

## Abstract

This research seeks to identify possible predictors of the difficulty level of reading comprehension items used in a standardized psychometric test for university admission. Several potential predictors of difficulty were proposed, namely, propositional density, negations, grammatical structure, vocabulary difficulty, presence of enhancement elements (words highlighted typographically), item abstraction level and degree of similarity between correct option and relevant text to resolve the item. By Linear Logistic Test Model (Fisher, 1973) it was found that the number of propositions, the syntactic structure, and fundamentally, the presence of difficult words contributed to the prediction of the item difficulty level.

## Keywords

Cognitive psychology, language processing, reading comprehension, Item Response Theory, Linear Logistic Test Model, Task Analysis, Item difficulty level.

## Resumen

Esta investigación busca identificar posibles variables predictoras del nivel de dificultad de los ítems de comprensión de lectura utilizados en una prueba psicométrica estandarizada para la admisión a una institución universitaria. Se propusieron varios posibles predictores del nivel de dificultad, a saber: densidad proposicional, negaciones, estructura sintáctica, dificultad del vocabulario, presencia elementos de realce (palabras resaltadas tipográficamente), abstracción del ítem y grado de similitud entre opción correcta y texto relevante para resolver el ítem. Mediante el Modelo Logístico Lineal de Rasgo Latente se encontró que la cantidad de proposiciones, la estructura sintáctica y, fundamentalmente, la presencia de léxico difícil de comprender contribuyeron a la predicción del nivel de dificultad.

## Descriptores

Psicología cognitiva, Procesamiento del lenguaje, Comprensión de lectura, Teoría de Respuesta al Ítem, Modelo Logístico Lineal de Rasgo Latente, Análisis de tareas, Nivel de dificultad de los ítems.

La investigación propuesta consiste en llevar a cabo un estudio de los ítems verbales de la Prueba de Aptitud Académica de la Universidad de Costa Rica (PAA), para lo cual se proponen un conjunto de atributos que podrían

aumentar o disminuir sus índices de dificultad. Para ello se utiliza el Modelo Logístico Lineal de Rasgo Latente (LLTM, por sus siglas en inglés), el cual permite relacionar estadísticamente los atributos propuestos para cada ítem

y las respuestas de los examinandos, ingresando estos datos en dos matrices y realizando los cálculos respectivos para descomponer el índice de dificultad de cada ítem (parámetro *b*) en los componentes involucrados en la resolución de cada uno.

Este proyecto sobre los ítems verbales de la PAA entronca con los enfoques más recientes sobre la validez (Gorin, 2007; Embretson, 1996; Embretson y Gorin, 2001; Messick, 1995; Cronbach y Meehl, 1955; Borsboom, Mellenbergh y Heerden, 2004), según los cuales la validez de las inferencias hechas a partir de las interpretaciones del resultado de un test está garantizada cuando a) el constructo existe como tal y b) las variaciones en el constructo causan variaciones en los puntajes obtenidos en el test. De esta manera, el objetivo primordial es identificar algunas características de los ítems verbales de la PAA que permitan predecir su nivel de dificultad, con base en algunas variables que han sido sumamente investigadas en el campo de la psicología cognitiva.

El interés por aislar las estrategias y atributos de los ítems involucrados en la resolución de test de comprensión de lectura es relativamente reciente, si se compara con la larga historia que poseen este tipo de pruebas (Hunt, 2011). Esta línea de investigación se ha construido a partir de la indagación (principalmente, mediante entrevistas semiestructuradas y autoreportes verbales) sobre las estrategias que utilizan los sujetos para resolver ítems de comprensión de lectura (Farr, Pritchard y Smitten, 1990; DeMauro, Merritty y Adams, 1994; Powers y Wilson, 1995; Rupp, Ferne y Choi, 2006). Asimismo, el intento de predecir el nivel de dificultad de los ítems ha generado una gran cantidad de investigaciones en las que el interés fundamental es identificar variables predictoras de índole lingüística y cognitiva (Drum, Calfee y Cook, 1981; Stahl y Jacobson, 1986; Davey, 1988; Kirsch y Mosenthal, 1990; Freedle y Kostin, 1991; Freedle y Kostin, 1992; Daneman y Hannon, 2001; Sheehan y Ginther, 2001; Gorin y Embretson, 2006; Ozuru, Rowe, O'Reilly y McNamara,

2008; Leroy, Helmreich y Cowie, 2010). En estas últimas, se ha utilizado un enfoque primordialmente cuantitativo, echando mano de técnicas estadísticas tradicionales como la regresión múltiple, el análisis de varianza y las correlaciones; sin embargo, también se encuentran diversos estudios en los que se utiliza el LLTM para estimar la capacidad predictiva de ciertas características de los ítems de comprensión de lectura y de razonamiento lógico (Embretson y Wetzel, 1987; Sheehan y Mislavy, 1990; Gorin, 2005; Sonnleitner, 2008; Attorresi, Picón, Abal, Aguerri y Galibert, 2009; Gilbert, Picón, Lozzia, Aguerri y Attorresi, 2010; Poinstingl, 2009).

Así las cosas, los estudios referidos anteriormente han resultado ser de vital importancia para identificar algunas variables que podrían ser de interés para predecir el índice de dificultad de los ítems verbales de la PAA. Tanto los diferentes estudios que han utilizado el LLTM para dicho propósito como los trabajos que utilizan técnicas clásicas de análisis de datos cuantitativos han servido para justificar las fuentes de dificultad que se expondrán a continuación. Además, los estudios centrados en las estrategias para resolver este tipo de ítems son relevantes para tomar en cuenta la gran importancia que tienen las opciones de respuesta y las preguntas a la hora de predecir el nivel de dificultad de una tarea verbal determinada.

A partir de la revisión bibliográfica que se llevó a cabo, fue posible identificar un conjunto de fuentes de dificultad consideradas como relevantes para la presente investigación. Más adelante se hará una exposición detallada de estas fuentes de dificultad, así como también la forma en que serán codificadas. En la siguiente sección se expone el modelo teórico que sustenta la presente investigación, en el cual se presentará la incorporación de los predictores identificados en la bibliografía.

Es importante subrayar el hecho de que la pertinencia de los atributos de los ítems fue considerada a la luz de la experiencia del in-

investigador de este proyecto, así como también de expertos en el campo de la lingüística y la psicología cognitiva. Ciertamente, sería posible incluir muchos otros elementos que podrían influir en el nivel de dificultad de los ítems; sin embargo, ha sido necesario llevar a cabo un filtrado con base en la experiencia acumulada que se tiene sobre las características particulares de la PAA. En este sentido, los ítems utilizados en esta prueba no presentan todas las características de otros test de razonamiento verbal y comprensión de lectura.

Tomando en consideración el objetivo de la presente investigación, ha sido necesario integrar en un solo marco conceptual los aportes teóricos de distintas disciplinas. Por lo tanto, a continuación se expone el interés de la psicometría por los aportes de la psicología cognitiva, las características del LLTM, el modelo de Embretson y Wetzel (1987) para explicar en términos cognitivos cómo se resuelve un ítem de comprensión de lectura y, finalmente, las fuentes de dificultad presentes en dicho proceso.

### **Psicometría y psicología cognitiva**

Actualmente, los modelos psicométricos de rasgo latente son de gran interés debido a que permiten llevar a cabo la calibración de los ítems de un test a partir de sus atributos. Esto es, mientras que en la Teoría Clásica de los Test solo era posible calcular medidas globales del test, estos enfoques relacionados con la medición de variables o rasgos latentes permiten una medición que cumple con el requisito de validez desarrollado al principio del presente documento.

Asimismo, la psicometría moderna ha mostrado un gran interés en utilizar los hallazgos de la psicología cognitiva para explicar teóricamente cuáles son los factores cognitivos que subyacen a los puntajes obtenidos en diferentes tipos de pruebas psicométricas (Cortada de Kohan, 2003). Esta posibilidad resulta de gran utilidad para la construcción de ítems con parámetros preestablecidos, así como también es de gran relevancia para explicar las caracterís-

ticas o componentes de los ítems que conforman los test psicométricos, tales como la PAA.

En este sentido, se parte del hecho de que los puntajes obtenidos en todo instrumento de papel y lápiz dependen, en cierta medida, de las características específicas que posee cada ítem. Por tal motivo, es esperable que el resultado obtenido en ítems verbales se relacione con aspectos propios de la estructura del ítem, esto es, con características asociadas a las lenguas naturales.

Por otro lado, el puntaje obtenido en un test psicométrico está asociado a un constructo denominado tradicionalmente como “habilidad”, la cual se conceptualiza como una variable dependiente a partir de los cálculos estadísticos realizados con los puntajes obtenidos en el test. En otras palabras, la habilidad se obtiene como un indicador sumario que involucra una combinación de procesos cognitivos, estrategias, componentes y conocimientos que pueden ser objeto de modelación matemática a través de una teoría de medición, en el caso particular de esta investigación, el modelo de Rasch y el LLTM (Embretson, 1989).

Ahora bien, no se debe olvidar que, aunque psicología cognitiva y psicometría converjan, el interés fundamental en el ámbito de investigación presentado anteriormente es la búsqueda de evidencias de validez y el mejoramiento del poder predictivo de los instrumentos psicométricos, para poder establecer diferencias más precisas entre los niveles de habilidad de los sujetos. Si se es capaz de identificar aquellos atributos de los ítems que explican su dificultad, se disfrutaría de la posibilidad de tener una mayor exactitud en la medición del constructo, lo cual permite llevar a cabo mediciones válidas y confiables y establecer diferencias igualmente pertinentes entre los sujetos que contestan un test. Por ello, a este tipo de modelos psicométricos también se les ha denominado *modelos explicativos de respuesta al ítem* (De Boeck y Wilson, 2004), porque permiten explicar en términos cognitivos las variaciones que muestran los estadísticos cal-

culados con base en las respuestas de los examinandos.

### El Modelo Logístico Lineal de Rasgo Latente (LLTM)

En la presente investigación se emplea uno de los modelos logísticos de rasgo latente que más se ha utilizado para modelar los componentes involucrados en la resolución de un ítem: el Modelo Logístico Lineal de Rasgo Latente (o LLTM, por sus siglas en inglés: *Linear Logistic Test Model*). Fue propuesto por Fisher (1973) y es considerado uno de los muchos modelos que se derivan del propuesto por Rasch (Bond y Fox (2001) brindan una presentación detallada de este modelo). De acuerdo con este modelo de medición, responder un ítem no es un fenómeno determinístico, sino probabilístico, de modo que contestarlo correctamente no se relaciona linealmente con la puntuación total del test. Si la probabilidad de acertar un ítem se calcula en cada nivel de habilidad de los examinados, obtenida a partir del puntaje total en el test, esta se distribuye normalmente, de modo que la curva resultante asume la forma de una normal acumulativa; y, además, si se toma en cuenta que una curva logística es muy similar en forma pero matemáticamente más simple, se verá por qué la siguiente ecuación propuesta por Rasch ha resultado ser de gran utilidad en el campo de la psicometría:

$$P(X_{ij}=1|\theta_j, b_i) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)}$$

De forma similar a como ocurre en una regresión logística, en la fórmula anterior  $b_i$  es la dificultad del ítem  $i$ ,  $\theta_j$  es la habilidad de la persona  $j$  y  $P(X_{ij} = 1)$  es la probabilidad de que la persona  $j$  acierte el ítem  $i$  (Embretson, 1989). De acuerdo con esta función planteada por Rasch (Martínez, Hernández y Hernández, 2006), la probabilidad de contestar correctamente un ítem está influenciada fundamentalmente por la habilidad del examinado y la dificultad del ítem.

Cuando se observa la fórmula anterior, se puede notar que la dificultad del ítem (el parámetro  $b$ ) constituye una medida agregada, es decir, se calcula tomando en consideración las respuestas de todos los examinandos. Por su parte, el LLTM lo que permite es descomponer este índice de dificultad, utilizando para ello la siguiente fórmula:

$$P(X_{ij}=1|\theta_i, q_j, \eta) = \frac{\exp(\theta_i - \sum_{k=1}^K q_{jk} \eta_k)}{1 + \exp(\theta_i - \sum_{k=1}^K q_{jk} \eta_k)}$$

Fundamentalmente, lo que el LLTM hace es reemplazar el parámetro  $b$  por la combinación lineal de los factores involucrados en la resolución del ítem (Chen, MacDonald y Leu, 2011, Baker, 1993). En este sentido, dicho modelo es análogo a crear un modelo de regresión de la dificultad de los ítems (variable dependiente) sobre los factores o fuentes de dificultad de estos (variables predictoras) (Embretson, 2010).

### Resolución de ítems verbales

A partir de la modelación matemática que permite el LLTM, es posible diseñar ítems que cumplan con los parámetros de dificultad establecidos a priori. Sin embargo, para identificar cuáles son esos factores o fuentes de dificultad presentes en los ítems, es imprescindible acudir a una teoría sustantiva pertinente que dé cuenta sobre cuáles son los atributos que influyen sobre la respuesta de los examinandos. Para este propósito, el modelo planteado por Embretson y Wetzel (1987) para la resolución de ítems de comprensión de lectura, resulta de gran interés.

Antes de explicar detalladamente este modelo, es pertinente recordar cuál es la estructura fundamental de un ítem verbal, entre los cuales existen dos formatos: de comprensión de lectura y de completar oraciones. En el caso de los ítems con el primer formato, primero se presenta un pasaje o texto, después del cual se brinda una instrucción o pregunta y a continuación una serie de opciones, una de las cua-



les constituye la respuesta correcta. El segundo formato es similar al primero, con la diferencia de que no se incluye una pregunta dentro de cada ítem, sino que se brinda una instrucción general al principio del apartado de completar oraciones, en el que se solicita al examinando que, para cada uno de los ítems, marque con equis la(s) palabra(s) que completan el sentido del pasaje. La extensión del texto puede variar, sin embargo, en la PAA a lo sumo alcanza una extensión de 12 renglones, aproximadamente unas 100 palabras. Cabe mencionar que los ítems con formato de completar oraciones suelen ser más cortos, dado que carecen de una pregunta entre el texto principal y las opciones de respuesta.

Ahora bien, la propuesta de Embretson y Wetzel para estudiar el procesamiento cognitivo de este tipo de ítems es un modelo compuesto por dos grandes etapas de procesamiento: 1) representación del texto principal o pasaje y 2) elección de la opción correcta. En la primera etapa ocurre la comprensión de lectura propiamente dicha, mientras que en la segunda operan diversos procesos relacionados con la comprensión de lectura y con el formato de pregunta de elección múltiple. A continuación, se expone con mayor detalle en qué consisten ambas etapas.

#### *Etapas 1: La representación del texto principal*

Esta etapa gira alrededor de la codificación léxica y la coherencia discursiva. Ambas están reguladas por las características lingüísticas del pasaje, especialmente por la posibilidad de entender el significado de las palabras utilizadas, de modo que el vocabulario con un mayor nivel de dificultad tenderá a hacer más difícil el proceso de comprender el texto. Por su parte, la coherencia discursiva se relaciona con el proceso de vincular los significados léxicos y las proposiciones en una representación significativa del texto.

Antes de continuar con esta exposición, es pertinente aclarar lo que se entiende en la presente investigación como proposición. Dicho concepto se refiere a un estado, evento o ac-

ción representados en el texto que pueden ser evaluados en términos de su valor de verdad con respecto al mundo real o a una construcción imaginaria (Graesser, Millis y Zwaan, 1997). Ahora bien, dichos estados, eventos o acciones presentados en el texto se integran, junto con el significado léxico de las palabras, en un todo coherente, de modo que, en palabras de Kintsch (1998), “la comprensión implica formar, a partir de rasgos perceptuales y conceptuales, unidades coherentes con características tipo Gestalt” (p. 93).

La etapa de representación constituye un proceso cíclico o iterativo, en el que las proposiciones (en este caso, las unidades de procesamiento) del texto son integradas en un conjunto coherente, junto con el conocimiento previo que posee el lector. Por ello, al modelo propuesto por Kintsch y van Dijk (1978) se le llama Modelo de Construcción-Integración, en el tanto en que la comprensión de lectura es el proceso de construcción de una representación mental a partir de la integración de múltiples proposiciones, de modo que la dificultad en esta etapa viene dada por la densidad proposicional del texto y por todos aquellos factores que dificultan el proceso de construcción-integración, como el uso de negaciones u oraciones sintácticamente complejas (oraciones subordinadas o en voz pasiva). Efectivamente, la razón entre el número de proposiciones y el tamaño total del texto constituye un factor de peso a la hora de crear una representación mental coherente, ya que la capacidad para procesar dichas unidades es limitada. Asimismo, si estas no son integradas correctamente, la información no será adecuada para utilizarla posteriormente en la resolución de un ítem (Gorin y Embretson, 2006).

#### *Etapas 2: La elección de la opción correcta*

Esta segunda etapa del modelo es denominada por Embretson y Wetzel (1987) como proceso de decisión, en virtud de que es cuando las personas toman una decisión sobre cuáles son las opciones de respuesta incorrectas y cuál correcta. En un primer momento, los subprocesos que ocurren en la etapa anterior (co-

dificación léxica y coherencia) también se dan, de la misma manera, con la pregunta y con las opciones de respuesta (en el caso del formato de completar oraciones, estos procesos solo ocurren en las opciones de respuesta). En otras palabras, es necesario que el examinando primero comprenda lo que se le solicita, junto con las alternativas que se le presentan, antes de proceder con los restantes subprocesos. Sin embargo, no es suficiente comprender el significado de la pregunta y las opciones, porque también es necesario elegir la alternativa correcta en función de la pregunta o instrucción del ítem.

El siguiente paso es el mapeo del texto, esto es, el proceso de relacionar las proposiciones de las preguntas y las opciones con la representación previamente construida e integrada a partir de las proposiciones del pasaje. El nivel de dificultad en el mapeo textual se ve influenciado por la cantidad de información necesaria del pasaje para contestar la pregunta o llevar a cabo la tarea solicitada. Lo anterior implica que no toda la información del pasaje es imprescindible para resolver el ítem, sino que existen segmentos de texto relevante o información clave para encontrar la respuesta correcta entre las opciones.

Finalmente, el último paso dentro de esta segunda etapa es evaluar el estatus de verdad de las alternativas de respuesta, la cual a su vez se lleva a cabo mediante dos actividades: la falsación y la confirmación. Este proceso de naturaleza dual se relaciona con la cantidad de información expuesta en el pasaje que puede ser utilizada para falsar o confirmar las opciones de respuesta. En este sentido, la dificultad del ítem disminuye en la medida en que las opciones incorrectas puedan ser falsadas por el texto (porque las contradice explícitamente) o la alternativa correcta pueda ser confirmada por el pasaje (porque la señala explícitamente).

Adicionalmente, Embretson y Wetzel (1987) propusieron otras variables que pueden afectar el nivel de dificultad del ítem. Una es el nivel del vocabulario empleado en las opciones de

respuesta: un distractor que presenta términos de difícil comprensión tiende a no ser considerado como un candidato plausible para ser la opción correcta, así como también quienes contestan una pregunta de opción múltiple con frecuencia no seleccionan la respuesta correcta cuando el vocabulario utilizado en esta es poco accesible. Es decir, el vocabulario de difícil comprensión actúa como un factor que aumenta la dificultad del ítem (cuando está presente en la alternativa correcta y ausente de los distractores) o que la disminuye (cuando está ausente de la opción correcta y presente en los distractores). Otra fuente de dificultad involucrada en la etapa de elección de la respuesta es la similitud entre las opciones y el texto relevante para contestar el ítem. El nivel más bajo de dificultad se presenta cuando la opción correcta es igual al texto pertinente para contestar el ítem, ya que no es necesario llevar a cabo ningún proceso de inferencia ni de reordenamiento para contestar. Por otro lado, el nivel más alto de dificultad sería aquel en donde es necesario llevar a cabo una generalización o una aplicación, las cuales ocurren cuando la alternativa correcta es un enunciado general que explica los casos particulares del pasaje o cuando dicha alternativa es un caso particular de alguna regla o enunciado general del pasaje, respectivamente. Entre ambos polos de dificultad, existe un nivel intermedio en el que la opción correcta es una paráfrasis de la porción de texto relevante para contestar el ítem, en otras palabras, la opción correcta expresa la misma información que el texto relevante para contestar el ítem pero con otras palabras (a través de sinónimos, reordenamientos sintácticos, frases equivalentes, etc.).

#### *Fuentes de dificultad de los ítems verbales de la PAA*

Una vez presentados todos los elementos que sustentan teóricamente esta investigación, solo resta proponer un listado de cuáles son las características o atributos que podrían explicar el nivel de dificultad de los ítems verbales de la PAA. A continuación, se presentan las características o atributos propuestos: Densidad proposicional predicativa y modificadora, canti-

dad de conectores, negaciones, oraciones pasivas con “se”, oraciones subordinadas, dificultad léxica, frecuencia de uso de vocabulario y presencia de elementos de realce (palabras escritas en negrita).

Cabe señalar que se utilizaron muchas de las fuentes de dificultad propuestas por Embretson y Wetzel (1987) y Gorin y Embretson (2006), quienes llevaron a cabo exhaustivas búsquedas bibliográficas para incorporar en sus análisis todas aquellas variables que influyen en la comprensión de lectura.

### Hipótesis

En este apartado, se explican las hipótesis derivadas del modelo de Embretson y Wetzel (1987), de los antecedentes de investigación y de la experiencia sobre las características particulares de la PAA.

*H<sub>1</sub>: A mayor densidad proposicional predicativa y modificadora, mayor es la dificultad a la hora de resolver el ítem.*

Se espera que una mayor cantidad de proposiciones signifiquen una mayor demanda de procesamiento para el examinando; lo anterior es esperable por cuanto este debe mantener activa una mayor cantidad de información en la medida en que el ítem presenta más proposiciones. A esta hipótesis corresponden las fuentes de dificultad referidas a la densidad proposicional predicativa y modificadora en el pasaje, la instrucción y las opciones de respuesta.

*H<sub>2</sub>: A mayor densidad proposicional conectiva, menor será el nivel de dificultad.*

Las partículas de enlace (palabras como “pero”, “además”, “en consecuencia”, “sin embargo”, etc.) facilitan la integración coherente del texto dándole al estudiante las pistas para identificar las relaciones lógicas y conceptuales entre las diferentes proposiciones del texto principal.

*H<sub>3</sub>: La presencia de negaciones dificulta la resolución de los ítems.*

Utilizar la partícula “no” para referirse a una proposición determinada obliga al lector a mantener en mente dos versiones de una misma situación: la que se refiere a la situación en sí misma y la que se relaciona con la ausencia o falsedad de lo dicho. En este sentido, cada proposición negada aporta mayor información que una proposición afirmativa. Por ejemplo, para comprender la proposición “Mañana no iré a clases” es necesario identificar el evento de “ir a clases” y además procesar la no ocurrencia de la misma acción. En cuanto a las fuentes de dificultad planteadas previamente, esta hipótesis se aplica al pasaje, a la pregunta (si la hubiere) y las alternativas de respuestas.

*H<sub>4</sub>: La presencia de oraciones pasivas y subordinadas incrementa la dificultad a la hora de resolver el ítem.*

En las investigaciones consultadas, este tipo de construcciones han sido consideradas como factores importantes en el aumento del nivel de dificultad a la hora de comprender un texto. Es esperable que toda oración pasiva deba ser transformada a la voz para ser incorporada en la red proposicional que se construye a partir de un texto. Por tal razón, se espera que (al igual que con las negaciones) tal transformación represente una mayor demanda para el lector. Asimismo, la subordinación implica una mayor cantidad de procesamiento en términos de memoria de trabajo, ya que demanda del lector mantener activo el sujeto de la oración mientras procesa el resto de constituyentes de la oración, los cuales suelen estar relativamente alejados de este precisamente por la estructura típica de una oración subordinada.

*H<sub>5</sub>: El vocabulario de difícil comprensión aumentará la dificultad de los ítems cuando esté presente en la opción correcta y en la medida en que haya este tipo de palabras en el texto principal.*

En la literatura consultada se encontró que quien contesta un ítem necesita conocer el significado de las palabras que lo componen. Cuando dicho conocimiento léxico no existe, la comprensión se empobrece en la medida en que la persona no puede utilizar toda la infor-

mación requerida para resolver un ítem. Particularmente, cuando uno o varios términos poco conocidos (es decir, de difícil comprensión) están en la respuesta correcta, esta tiende a ser considerada como poco plausible, de modo que los distractores con léxico más accesible se vuelven mejores candidatos (desde la perspectiva de quien contesta) a ser la opción correcta.

*H<sub>6</sub>: La presencia de elementos de realce (palabras en negrita) facilita la resolución de un ítem.*

Esto se explica porque este recurso tipográfico facilita visualmente al examinando determinar qué es exactamente lo que debe buscar en las opciones de respuesta.

## **Método**

### **Muestra**

La muestra de ítems está conformada por 100 ítems verbales aplicados en el año 2011, los cuales son utilizados regularmente en el examen de admisión de la Universidad de Costa Rica. Cabe señalar que los ítems empleados para el análisis serán solamente los de banco, es decir, los que en ese año cumplían con todos los requisitos psicométricos que debe poseer un ítem de aplicación regular. Dado lo anterior, en el análisis no se tomaron en cuenta los ítems experimentales, es decir, aquellos que por primera vez se aplicaron en ese año, por cuanto en ese momento no se tenía certeza de que fueran representativos de la población de ítems verbales de la PAA. Por otro lado, 37937 estudiantes de secundaria contestaron la PAA en ese año.

### **Procedimiento**

Tres jueces expertos y un codificador juzgaron los ítems que componen la muestra: el codificador se encargó de codificar los 100 ítems de acuerdo con los atributos en los que solamente se requerían conteos, mientras que los jueces se encargaron de los predictores en los cuales se debía realizar una valoración más global. Una vez finalizada esta labor, se calcularon los respectivos índices de concordancia (kappa de Cohen y correlación intraclase).

Posteriormente, se construyeron dos matrices de datos: una con las respuestas de los y las estudiantes a los ítems (en términos de 1 = acierto y 0 = fallo) y otra con los valores asignados por los jueces expertos y el codificador a los mismos ítems. Dichas matrices fueron ingresadas en el software R con el fin de verificar la correlación existente entre el índice de dificultad estimado por este modelo y el estimado mediante el modelo de Rasch. Finalmente, mediante un modelo de regresión se identificaron cuáles eran las fuentes de dificultad estadísticamente significativas que mejor predicaban el parámetro b (índice de dificultad en el modelo de Rasch).

Cabe aclarar que los jueces expertos fueron elegidos con base en sus respectivas áreas de experticia, a saber: lingüística (dos expertos) y psicología cognitiva (un experto). Asimismo, se tomó en cuenta la restricción que existe en cuanto al acceso a los ítems, ya que estos son material confidencial del Programa Permanente de la Prueba de Aptitud Académica, por lo cual se debió escoger a profesionales que pudieron garantizar la discrecionalidad en el uso de la información que iban a manipular. En este sentido, estas personas firmaron un compromiso de confidencialidad con el Programa mediante el cual se comprometieron a no divulgar dicha información.

Originalmente se plantearon más predictores de la dificultad, no obstante, durante la codificación fue posible comprobar que algunos no eran relevantes o no estaban presentes en la muestra de ítems, por lo que se decidió excluirlos de los análisis. Por otro lado, para medir la frecuencia de uso léxica se utilizaron dos corpus elaborados por investigadores del programa *Estudios de Lexicografía Hispano-Costarricense*, del Instituto de Investigaciones Lingüísticas de la Universidad de Costa Rica (Sánchez, s.f.; Ríos, 2011).

Para llevar a cabo los análisis, se utilizaron dos programas informáticos: R para ejecutar el LLTM y el modelo de regresión, y Winsteps para correr el modelo de Rasch. Asimismo, se



elaboró un manual de codificación y una hoja de registro para que los jueces expertos analizaran los ítems a partir de las fuentes de dificultad presentadas anteriormente.

### **Estrategia de análisis**

Debido a la índole fundamentalmente psicométrica de este proyecto, la estrategia de análisis constituye un componente intrínseco de la teoría que lo sustenta. Por tal motivo, en el marco teórico se explicó detalladamente en qué consiste el LLTM, de modo que en este apartado solamente se mencionarán los pasos requeridos para llevar a cabo el análisis.

Antes de exponer los pasos para implementar el LLTM, es importante señalar que solamente se trabajó con aquellos ítems que cumplieran con los supuestos del modelo de Rasch (unidimensionalidad de los ítems y ajuste al modelo). Por ello, en primer lugar se ejecutó un análisis factorial exploratorio y, mediante el programa Winsteps, se utilizó el modelo de Rasch para evaluar el ajuste de los ítems y de las personas.

Así las cosas, a continuación se enumeran las etapas del LLTM:

- 1.- Se ejecutó el modelo de Rasch en R para estimar los índices de dificultad de los ítems. Este análisis se hizo por formulario, de tal manera que se hicieron cuatro análisis separados, de conformidad con la manera en que se aplica la Prueba de Aptitud Académica.
- 2.- Los códigos asignados por los jueces expertos se ingresaron, como una matriz, al eRm (paquete implementado en R para ejecutar el modelo de Rasch y el LLTM).

3.- Se utilizó el LLTM para estimar los índices de dificultad con base en los predictores ingresados en el paso dos.

4.- Se evaluó el ajuste del modelo mediante los coeficientes de correlación entre las estimaciones mediante el modelo de Rasch y el LLTM.

5.- Para identificar la dirección e importancia del efecto específico para cada predictor, se utilizó un modelo de regresión múltiple.

### **Resultados**

A continuación se exponen los resultados de los análisis realizados. Este apartado está dividido en siete secciones, de acuerdo con la información proporcionada en cada una, a saber: 1) estadísticas descriptivas, 2) coeficientes de concordancia entre jueces, 3) estructura factorial de cada formulario, 4) ajuste de los ítems al modelo de Rasch, 5) índices de bondad de ajuste del LLTM, 6) parámetros eta del LLTM y 7) coeficientes beta estandarizados del modelo de regresión.

#### **Estadísticos descriptivos**

En la Tabla 1 se presentan algunos datos descriptivos sobre los cuatro formularios de la PAA. En primer lugar, se presenta la cantidad de personas que contestó cada uno, la cual representa una cantidad sumamente grande en comparación con las muestras que suelen aparecer en la investigación psicométrica. Los formularios son lo que se denomina pruebas paralelas, con pequeñas diferencias del orden de los decimales entre cada uno. La confiabilidad es alta y homogénea en todos los formularios. Finalmente, en cuanto a la dificultad estimada en Rasch, cabe señalar que los ítems verbales en promedio presentan dificultades medias, tomando en cuenta que el rango suele ser de -3 a +3.

Tabla 1. Estadísticos descriptivos

	Formulario 1	Formulario 2	Formulario 3	Formulario 4
N	9480	9447	9418	9492
Media	18.38	19.18	19.12	18.47
Desviación estándar	7.84	7.72	7.89	7.78
Dificultad clásica	.46	.48	.48	.46
Discriminación clásica	.35	.34	.36	.35
Alfa de Cronbach	.87	.86	.87	.87
Dificultad Rasch	0	0	0	0
Confiabilidad de los ítems	1	1	1	1
Confiabilidad de las personas	.86	.85	.86	.85

### Concordancia entre jueces

En una hoja de registro para cada ítem, a los jueces se les proporcionó un conjunto de 5 criterios que debían utilizar para evaluar los ítems, a saber: nivel de abstracción del ítem, similitud entre opción correcta y texto relevante para contestar el ítem, dificultad léxica del pasaje (incluía la instrucción), dificultad léxica de los distractores y dificultad léxica de la opción correcta. Estos aspectos, junto con las demás variables predictoras, fueron incluidos con base en las investigaciones hechas sobre la resolución de ítems verbales. No obstante, como se observa en la Tabla 2, los coeficientes de concordancia (kappa de Cohen entre pares de jueces, kappa para los tres jueces en conjunto y la correlación intraclase) fueron sumamente bajos en todos los casos (López y Pita, 1999; Clark-Carter, 2004).

Debido al nivel tan bajo de concordancia entre los jueces, se tomaron dos decisiones. En primer lugar, se descartó de los análisis posteriores el nivel de abstracción del ítem y la similitud entre el texto principal y la opción correcta del ítem. En segundo lugar, en lo que respecta a los tres criterios sobre dificultad léxica, fueron utilizados de manera independiente para cada uno de los jueces.

La primera decisión se justifica porque durante el proceso de juzgamiento, los jueces expresaron sus dudas sobre la pertinencia de dichos criterios. Tanto ellos como el investigador a cargo examinaron con mayor detenimiento los ítems y se llegó a la conclusión de

que la abstracción del ítem y la similitud entre la opción correcta y el texto no eran criterios adecuados a las características particulares de estos.

En cuanto a la segunda decisión, era esperable que hubiera un bajo nivel de concordancia, ya que sin un corpus léxico de referencia no es posible identificar objetivamente cuándo una palabra podría resultar desconocida para la población que anualmente contesta los ítems de la PAA. Por otro lado, no se excluyeron dichos criterios de los análisis posteriores porque sí son relevantes e importantes en la predicción de la dificultad. Además, en previsión de esta situación, se incluyó una variable de frecuencia de uso léxico, de modo que fuera posible determinar con un mayor nivel de objetividad cómo el léxico afecta la dificultad de los ítems.

Finalmente, en los estudios de Embretson y Wetzel (1987) y Gorin y Embretson (2006) se encontraron problemas similares en lo que respecta al acuerdo entre jueces. En este sentido, Gorin enfatiza la importancia de las características particulares de cada prueba, lo cual hace difícil identificar a priori atributos o características en un instrumento psicométrico que previamente hayan sido extraídos de otras pruebas. Así pues, dado que esta es la primera descripción detallada que se hace sobre los ítems verbales de la PAA, no fue posible prever que ciertos predictores identificados en pruebas de comprensión de lectura no son relevantes para caracterizar la PAA.

Tabla 2. Concordancia entre jueces

	Kappa para los tres jueces en conjunto	J1 – J2	J2 – J3	J1 – J3
Nivel de abstracción del ítem	.30	.45	.34	.19
Similitud entre opción correcta y texto relevante	.07	.15	.07	.10
Dificultad léxica del pasaje	.55*			
Dificultad léxica de los distractores	0	.18	.01	.04
Dificultad léxica de la opción correcta	.50	.48	.09	.29

\*Coeficiente de correlación intraclase. Entre paréntesis se muestra el kappa para los tres jueces en conjunto.

### Estructura factorial de los formularios de examen

Con el objetivo de corroborar si los ítems de cada formulario son unidimensionales, se ejecutó un análisis factorial exploratorio (con rotación Varimax) para cada uno.

En la Tabla 3 se presentan los resultados para cada formulario. Como se puede observar, el coeficiente KMO y el estadístico chi cuadrado (entre paréntesis se presentan los grados

de libertad) alcanzaron valores sumamente elevados, lo cual indica que las correlaciones entre los ítems son lo suficientemente altas como para identificar al menos un factor común. Por otro lado, en lo que respecta al supuesto de unidimensionalidad, se observa que el porcentaje de varianza explicada del primer factor extraído es mucho mayor que el del segundo factor, lo cual es evidencia a favor de que los ítems de cada formulario de examen están midiendo un mismo constructo.

Tabla 3. Análisis factorial exploratorio

	F1	F2	F3	F4
KMO	.97	.97	.97	.97
$\chi^2$ de Bartlet	44680.98(780)*	42069.703(780)*	46040.451(780)*	42942.873(780)*
Varianza 1° factor	17.15%	16.56%	17.42%	16.71%
Varianza 2° factor	2.92%	2.93%	2.96%	2.86%

\* $p < .001$

### Ajuste de los ítems al modelo de Rasch

Una vez confirmado el supuesto de unidimensionalidad, se procedió a utilizar el modelo de Rasch para identificar las personas que no se ajustaran al modelo. Este filtrado se realiza con base en un coeficiente denominado Infit, el cual debe ser un valor entre 0.7 y 1.3 de acuerdo con los lineamientos de Bond y Fox (2001). Así pues, del total de sujetos que contestaron la PAA, se eliminó a dos personas que respondieron el formulario 1, dos del formulario 2, siete del formulario 3 y un sujeto del formulario 4.

Posteriormente, se siguió este mismo criterio para eliminar los ítems que no se ajustaran. Para ello, se volvió a ejecutar el análisis únicamente con las personas cuyo Infit fue mayor a 0.7 y menor a 1.3. No obstante, en esta se-

gunda ejecución del modelo no fue necesario eliminar ningún ítem, debido a que todos mostraron un Infit mayor de 0.8 y menor a 1.2, además de un PTME mayor que 0.10 (Bond y Fox, 2001). Así las cosas, una vez garantizado el cumplimiento de los supuestos sobre unidimensionalidad y ajuste al modelo de Rasch, se procedió a ejecutar el modelo LLTM para poner a prueba el poder explicativo de las variables predictoras.

### Bondad de ajuste de los modelos LLTM

Tal y como se mencionó en el apartado sobre la estrategia de análisis, estos se ejecutaron por formulario (F1, F2, F3 y F4) y por juez (J1, J2 y J3). Para llevar a cabo los análisis, en cada formulario de examen se siguió el siguiente procedimiento: primero se ejecutó el LLTM sin los atributos evaluados por los jueces ex-

peritos (dificultad léxica), en una segunda fase se incorporaron los valores asignados por el juez 1 a las dificultades léxicas, en la tercera se excluyeron los datos del juez 1 y se incluyeron los del juez 2 y en una cuarta ejecución del modelo se retiraron los datos del juez 2 para incorporar los del juez 3. De esta forma, para cada uno de los cuatro formularios se crearon cuatro modelos (MSJ, MJ1, MJ2 y MJ3), de los cuales se reportan tres índices de bondad de ajuste (Real, Olea, Ponsoda, Revuelta y Abad, 1999): el chi cuadrado de Andersen (Tabachnick y Fidell, 2007), la correlación de Pearson entre los índices de dificultad estimados en el modelo de Rasch y los correspondientes al LLTM, y el porcentaje de varianza explicado (ver Tabla 4). Como se puede inferir de lo anterior, los modelos MJ1, MJ2 y MJ3 corresponden al atributo *dificultad léxica* evaluado por cada juez. Por su parte, en el modelo cero (MSJ) se utilizaron los dos corpus léxicos para codificar la frecuencia de uso de las palabras presentes en el ítem.

Con la prueba de Andersen se busca que el chi cuadrado no sea significativo, ya que este estadístico señala la existencia de diferencias entre los índices de dificultad estimados con el LLTM respecto de los estimados con el modelo de Rasch. Sin embargo, dicha prueba es

muy sensible al tamaño de la muestra, por lo que es usual que resulte estadísticamente significativa. Por tal motivo, no es conveniente tomarla en cuenta como criterio decisivo para evaluar el ajuste del modelo LLTM. Inclusive, el mismo creador del modelo LLTM (Fischer, 1997) plantea la dificultad de obtener valores no significativos en esta prueba. Dados esos problemas, también se puede utilizar la correlación de Pearson para verificar si los parámetros estimados con el LLTM reproducen los índices de dificultad del modelo de Rasch (Romero, Ponsoda y Ximénez, 2008). Además, cuando se eleva al cuadrado, dicha correlación es una aproximación a la proporción de la varianza de los índices de dificultad en Rasch explicada por los parámetros del LLTM (Revuelta y Ponsoda, 1998).

Lo primero que salta a la vista en la Tabla 4 son las magnitudes de las chi cuadrado, todas ellas significativas. Sin embargo, tomando en cuenta que para cada formulario la muestra es de aproximadamente 9000 personas, no es de extrañar que estos valores sean tan grandes y, por su puesto, estadísticamente significativos. Sin embargo, las correlaciones de Pearson alcanzaron valores entre moderados y altos (la más baja fue de 0.71), junto con los respectivos porcentajes de varianza explicada.

Tabla 4. Índices de bondad de ajuste

		MSJ	MJ1	MJ2	MJ3
F1	$\chi^2$	7565(14)	8939(17)	7327(17)	10492(17)
	$r$	.86	.83	.86	.80
	$r^2$	74%	69%	75%	63%
F2	$\chi^2$	4343(16)	7237(19)	7498(19)	7339(19)
	$r$	.92	.86	.86	.86
	$r^2$	85%	74%	74%	75%
F3	$\chi^2$	14333(16)	12736(19)	14828(19)	15280(19)
	$r$	.73	.77	.73	.71
	$r^2$	54%	59%	53%	51%
F4	$\chi^2$	5475(14)	5754(17)	6576(17)	6392(17)
	$r$	.88	.87	.86	.86
	$r^2$	78%	76%	73%	74%

### Parámetros eta de cada modelo

Antes de iniciar la exposición de los resultados en cuanto los parámetros eta de cada atributo, es importante mencionar que fue necesaria

eliminar dos atributos del modelo LLTM en los formularios 2 y 3: negaciones en la instrucción y palabras en el pasaje. Respecto del número de negaciones, estos formularios so-



lamente un ítem presentaba esta característica, por lo cual el software no era capaz de realizar las estimaciones correspondientes debido a la falta de variabilidad. Por otro lado, se presentó el caso de un ítem con 116 palabras en el texto principal, un valor extremo que impedía la realización de los cálculos.

En la Tabla 5 se presentan estos parámetros por formulario; aquellos atributos que no resultaron ser significativos al 5% se señalan con negrita. Además, se utilizan las siguientes abreviaturas para facilitar la legibilidad del texto: P (pasaje o texto principal), I (instrucción), O (opciones de respuesta), PI (pasaje e instrucción), C (clave o respuesta correcta), D (distractores), A (corpus léxico uno) y B (corpus léxico dos). Finalmente, hay cuatro espacios en blanco correspondientes a la exclusión de dos atributos en la estimación del LLTM para los formularios 2 y 3. En esta tabla no se presentan los atributos de dificultad léxica evaluados por los jueces, los cuales se colocaron aparte en aras de facilitar la presentación, ya que para cada juez se requirió la estimación de un modelo LLTM en cada formulario. Cabe señalar que el software R estima las “facilidades” de los ítems, por lo cual a los parámetros  $\eta$  del modelo LLTM se les cambió manualmente el signo para poder interpretarlos en términos de la dificultad (Mair y Hatzinger, 2012).

Los atributos que no afectan significativamente la dificultad a la hora de resolver los ítems en al menos uno de los formularios son densidad adverbial del texto principal, la densidad adjetival de la instrucción, las negaciones de las opciones de respuesta, la no aparición en el corpus A (Sánchez, s.f.) de una o más palabras presentes en los distractores, la no aparición en el corpus B (Ríos, 2011) de vocablos presentes en el pasaje, la instrucción

y la opción correcta; y la cantidad de palabras en la instrucción. Dado que el valor  $p$  de estos atributos es mayor al 5%, se puede concluir que en algunos formularios no influyen en la dificultad de los ítems.

En la Tabla 6 se presentan los resultados correspondientes a los atributos evaluados por los jueces. Los que no son significativos en cuanto a la predicción de la dificultad son la dificultad léxica de los distractores para el juez 1 y 2, y la del pasaje y la instrucción en el caso del juez 2, en los casos específicos de los formularios 2 y 3.

Los modelos planteados se ajustaron muy bien a los datos (tomando en cuenta que el porcentaje más pequeño de varianza explicada es de 51%), lo cual indica que las variables elegidas fueron relevantes respecto del objetivo de la presente investigación, a saber: identificar algunos atributos de los ítems verbales de la PAA mediante los cuales se pueda predecir su nivel de dificultad. De hecho, la correlación más baja fue de  $r = .71$  entre los índices de dificultad calculados mediante el modelo de Rasch (Beta RM) y los correspondientes al modelo LLTM (Beta LLTM). Se puede observar que en la gran mayoría de los ítems se logró un grado satisfactorio de correlación, ya que son pocos los que se alejan de la línea diagonal.

Ahora bien, con el modelo LLTM fue posible identificar que la mayoría de los predictores propuestos tiene un efecto estadísticamente significativo sobre la dificultad de los ítems. En otras palabras, se puede afirmar con certeza que los resultados observados no son producto del azar, sino que pueden ser generalizados a la población –que, en el caso de la presente investigación, sería la población de ítems del banco verbal de la PAA.

Tabla 5. Parámetros eta de atributos codificados

	F1	F2	F3	F4
Pasivas con “se”	0.06	-0.20	0.20	-0.06
Oraciones subordinadas	0.06	0.19	0.16	0.02
Conectores	0.05	-0.25	0.11	0.01
Densidad adverbial P	<b>-0.00</b>	0.36	0.25	0.37
Densidad adverbial I	1.35	0.28	0.92	-1.45
Densidad adverbial O	0.12	-0.10	-0.14	0.03
Densidad adjetival P	-0.02	0.10	0.03	-0.05
Densidad adjetival I	0.64	0.06	<b>0.00</b>	0.14
Densidad adjetival O	0.02	-0.04	-0.02	-0.06
Densidad verbal P	-0.13	-0.15	-0.14	-0.24
Densidad verbal I	-0.11	-0.66	-0.39	-0.81
Densidad verbal O	0.10	0.04	-0.08	0.04
Negaciones P	-0.37	0.21	0.12	0.08
Negaciones I	-1.22			-0.80
Negaciones O	-0.16	-0.11	-0.16	<b>0.00</b>
No aparición PI A	-0.28	-0.13	0.06	-0.05
No aparición C A	-0.17	0.51	-0.44	0.18
No aparición D A	0.17	<b>0.01</b>	<b>0.02</b>	<b>-0.01</b>
No aparición PI B	-0.05	<b>0.00</b>	-0.06	-0.11
No aparición C B	0.26	-0.49	0.21	<b>-0.02</b>
No aparición D B	0.17	-0.07	-0.28	0.10
Elementos de realce	-0.32	0.31	-0.32	-1.05
Palabras P	0.03			0.04
Palabras I	-0.03	0.07	<b>0.00</b>	0.01
Palabras O	0.03	-0.03	-0.003	-0.01

Tabla 6. Parámetros eta de dificultad léxica

		F1	F2	F3	F4
J1	Dificultad léxica PI	0.07	-0.04	0.31	0.04
	Dificultad léxica D	0.27	<b>0.01</b>	0.19	0.26
	Dificultad léxica C	0.61	0.23	0.48	0.29
J2	Dificultad léxica PI	-0.10	<b>-0.01</b>	0.08	-0.10
	Dificultad léxica D	-0.78	0.08	<b>0.01</b>	-0.08
	Dificultad léxica C	0.22	0.12	0.37	0.17
J3	Dificultad léxica PI	-0.26	-0.18	-0.12	-0.17
	Dificultad léxica D	-0.42	0.43	-0.61	-0.46
	Dificultad léxica C	-0.17	-0.07	-0.21	-0.24

### Coefficientes beta estandarizados

Ciertamente, lo anterior constituye un aporte valioso; sin embargo, resulta de gran interés (de acuerdo con los objetivos específicos de este proyecto) identificar la importancia práctica de cada predictor, su efecto específico sobre la dificultad (si la disminuye o la aumenta) y la utilidad de cada uno respecto de los demás, es decir, cuáles atributos tienen un mayor poder predictivo. Debido a que las variables ingresadas en el modelo LLTM no presentan la misma escala ni rango, no es posible comparar entre sí los atributos de la Tabla 5 ni los de la Tabla 6 para identificar cuáles atribu-

tos predicen mejor la dificultad y en qué dirección.

Así las cosas, se ejecutó un modelo de regresión múltiple en el cual la variable dependiente es el índice de dificultad estimado mediante el modelo de Rasch y las independientes, los atributos de los ítems. De esta manera, se pueden calcular los coeficientes beta estandarizados, lo cual permite comparar los atributos entre sí para averiguar cuáles tienen un mayor poder predictivo respecto de la dificultad de los ítems.

La diferencia crucial entre el LLTM y un modelo de regresión es que aquel utiliza la información de todos los sujetos que contestaron los ítems, mientras que en una regresión solamente se cuenta con el índice de dificultad de cada ítem. Así pues, mientras que los parámetros eta del LLTM se estiman con base en la información aportada por aproximadamente 9000 sujetos, los coeficientes de regresión se calculan tomando en cuenta únicamente los 40 ítems de cada formulario. Lo anterior implica que los errores estándar son mucho mayores en el modelo de regresión y, por consiguiente, los valores p (significancia estadística) superan por mucho el criterio convencional de .05. Sin embargo, mediante el LLTM fue posible establecer que casi todos los predictores ejercen un efecto significativo ( $p < .05$ ) sobre la dificultad de los ítems, por lo cual los valores p resultantes del modelo de regresión no serán tomados en cuenta como criterio, sino solamente el signo y la magnitud de los coeficientes beta estandarizados.

Para el modelo de regresión, fue necesario eliminar y unificar ciertas variables debido a la asimetría que muestran y la correspondiente falta de variabilidad. En este sentido, en el

modelo de regresión no fueron incluidas las siguientes variables: oraciones pasivas con se, elementos de realce y densidad proposicional. En lugar de esta última variable, se creó otra denominada *Proposiciones* que es la suma de adverbios, adjetivos y verbos para cada ítem. Por otro lado, se verificó que los atributos no presentaran problemas de colinealidad mediante el estadístico *FIV* (Factor de inflación de la varianza), cuyo valor no fue mayor a 3 en los atributos incluidos en el modelo (Tabachnick y Fidell, 2007).

Dado que el acuerdo entre jueces fue sumamente bajo, se procedió a crear un modelo para cada juez en los cuatro formularios de la PAA. Como se puede observar en la Tabla 7, el modelo con el juez 1 es el que presenta mayores valores de varianza explicada mientras que el juez 3 presenta los valores más bajos. Por tal razón, se decidió elegir el modelo con el juez 1 para interpretar la importancia práctica de los coeficientes de regresión estandarizados. Antes de proceder a la interpretación de estos, es importante señalar que la varianza explicada del modelo propuesto no es alta, lo cual se comentará en el apartado de conclusiones.

Tabla 7. Varianza explicada de modelos por juez y por formulario

Formulario	J1				J2				J3			
	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4
R <sup>2</sup>	.38*	.36*	.45*	.34*	.31	.31	.35*	.29	.23	.11	.20	.20

\* $p < .05$

En la Tabla 8 se presentan los pesos de regresión del modelo con la codificación de la dificultad léxica hecha por el juez 1. Antes de proceder a interpretar estos resultados a la luz de las hipótesis y los objetivos de esta investigación, cabe recalcar que los coeficientes estandarizados cambian de dirección y magnitud

en cada formulario de examen. Lo anterior se debe a que la PAA no se ensambla utilizando como criterio los atributos planteados en este trabajo, por lo cual es esperable que en cada formulario de examen se presenten estas variaciones.

Tabla 8. Coeficientes beta estandarizados

	F1	F2	F3	F4
Oraciones subordinadas	.07	.27	.14	.15
Conectores	.05	-.06	-.12	-.03
Palabras ausentes de corpus	-.02	-.12	-.06	-.21
Proposiciones	.07	-.13	-.02	.01
Palabras difíciles en pasaje e instrucción	.30	.53	.58	.49
Palabras difíciles en distractores	-.15	-.21	-.14	-.17
Palabras difíciles en respuesta correcta	.43	.24	.24	.27

\*Variable dependiente: Dificultad estimada mediante el modelo de Rasch

Atendiendo a las hipótesis planteadas en el apartado respectivo, en relación con la primera hipótesis planteada ( $H_1$ ) sobre la cantidad de proposiciones, no se puede concluir que una mayor cantidad de estas aumente de manera importante la dificultad de los ítems. Como se observa en la Tabla 8, en tres de los cuatro formularios el coeficiente es mucho menor a .10 (valor considerado como mínimo necesario para concluir que una variable independiente sea importante en la predicción de una dependiente). Lo mismo se puede aducir respecto de  $H_2$  (la densidad conectiva disminuye la dificultad), ya que aunque los coeficientes muestran la dirección esperada (el signo negativo), en tres formularios de examen el beta estandarizado de la densidad conectiva está muy por debajo de .10.

En lo que respecta a  $H_3$  (la presencia de negaciones aumenta la dificultad), no fue posible estimar su importancia práctica ya que resultó ser una variable muy asimétrica. Así, la cantidad de ítems que presentan negaciones es muy baja, lo cual provoca que la variabilidad se reduzca a niveles inadecuados para estimar un modelo de regresión.

Por su parte, se obtuvieron evidencias a favor de  $H_4$  (influencia de la estructura sintáctica) en relación con las oraciones subordinadas. En este sentido, en tres de los cuatro formularios los betas estandarizados son mayores a .10. Lo anterior indica que, de acuerdo con lo hallado en otras investigaciones, las oraciones subordinadas aumentan el nivel de dificultad a la hora de realizar tareas de comprensión verbal.

Ahora bien, en cuanto a  $H_5$ , los resultados son sumamente reveladores de cuál es uno de los factores más importantes en cuanto la resolución de los ítems verbales de la PAA. En la Tabla 8 se ve claramente que los coeficientes estandarizados de la dificultad léxica (en el pasaje y la instrucción, en los distractores y en la respuesta correcta) son mucho más grandes y consistentes que el resto. Como se planteó previamente, la presencia de palabras difíciles de comprender en los distractores facilita el ítem, mientras que si estas están en el pasaje, la instrucción o en la respuesta correcta, el nivel de dificultad aumenta. Por otra parte, la variable relacionada con la frecuencia de uso léxica no resultó ser importante porque solamente en uno de los cuatro formularios alcanza un valor mayor a .10. Lo que sí llama la atención es que en todos los formularios el coeficiente sea negativo, porque esto indicaría que la presencia de palabras difíciles en un ítem disminuye el nivel de dificultad. En las conclusiones se abordará esta situación y las posibles explicaciones a este hallazgo.

Finalmente, en cuanto a  $H_6$  (efecto de los elementos de realce) no fue posible incorporar la variable respectiva al modelo de regresión porque presentó una distribución muy asimétrica, hecho que ocurre porque en los formularios elegidos para esta investigación hay muy pocos ítems que tengan palabras en negrita.

## Discusión

En este apartado se retomarán las hipótesis planteadas en el apartado correspondiente y se expondrá si estas se cumplieron o no, así como algunos comentarios explicativos. Posteriormente, se explicarán algunas medidas que a



futuro se podrían implementar para recabar más evidencias sobre el constructo medido por los ítems verbales de la PAA.

La primera hipótesis proponía que a mayor densidad proposicional predicativa y modificadora, mayor es la dificultad a la hora de resolver el ítem. Al respecto se ha planteado que una mayor cantidad de proposiciones implican una mayor demanda de procesamiento para el examinando, ya que este debe mantener activa una mayor cantidad de información en la medida en que el ítem presenta más proposiciones. En la presente investigación, no se podría afirmar que exista evidencia para sostener esta hipótesis, ya que la densidad proposicional tiene un efecto muy bajo en la dificultad de los ítems.

En lo que respecta a la segunda hipótesis (a mayor densidad conectiva, menor nivel de dificultad) podría sostenerse que un aumento en la cantidad de conectores implica una menor dificultad. En principio, las partículas de enlace (“pero”, “además”, “en consecuencia”, “sin embargo”, etc.) facilitan la integración coherente del texto porque le dan al estudiante las pistas para identificar las relaciones lógicas y conceptuales entre las diferentes proposiciones del ítem. En el presente trabajo, una mayor cantidad de conectores resultó estar asociada a menores niveles de dificultad.

Dado que los atributos anteriores están directamente relacionados con la carga informativa de los ítems, es plausible proponer que la poca variabilidad de estos haya afectado la estimación de los parámetros eta. En su gran mayoría, los ítems son sumamente cortos en comparación con los ítems tradicionalmente utilizados en pruebas de comprensión de lectura, en las cuales se presenta un texto largo (de hasta tres o cuatro párrafos) y varias preguntas asociadas a este. Por otro lado, la teoría utilizada en esta investigación (Kintsch, 1998; Embretson y Wetzel, 1987) ha sido utilizada para explicar el procesamiento de textos extensos y de uso cotidiano (noticias de periódico, cuentos, manuales, etc.), de modo que los atributos aso-

ciados a la densidad proposicional quizá hubieran resultado ser más contundentes en cuanto a su predicción si los ítems verbales de la PAA hubieran sido más variables y de mayor extensión (Gorin y Embretson (2006).

Retomando las hipótesis, en la tercera se planteó que la presencia de negaciones dificultaría la resolución de los ítems. El uso de la partícula “no” para referirse a una proposición determinada obliga al lector a mantener en mente dos versiones de una misma situación: la que se refiere a la situación en sí misma y la que se relaciona con la ausencia o falsedad de lo dicho. En este sentido, cada proposición negada aporta mayor información que una proposición afirmativa. No obstante, los ítems incluidos en la muestra (y en general los pertenecientes al banco verbal de la PAA) carecen en su gran mayoría de negaciones por lo cual no fue posible estimar su importancia práctica. Para futuras investigaciones de corte experimental se podrían construir ítems con negaciones para poner a prueba nuevamente esta hipótesis.

En cuanto a la cuarta hipótesis sobre la presencia de oraciones pasivas y subordinadas como factores de incremento de la dificultad, se obtuvo evidencia a favor de lo predicho sobre las oraciones subordinadas pero no sobre las oraciones pasivas. Era esperable que toda oración pasiva tuviera que ser transformada a la voz activa para ser incorporada en la red proposicional que se construye a partir de un texto. Por tal razón, se esperaba que tal transformación representara una mayor demanda cognitiva, pero en el caso del español al ser construcciones medio pasivas, es decir, con el verbo activo y el agente marcado con la partícula “se”, la situación es diferente. Por otro lado, la subordinación implica una mayor cantidad de procesamiento en términos de memoria de trabajo, ya que demanda del lector mantener activo el sujeto de la oración mientras procesa el resto de constituyentes de la oración, los cuales suelen estar relativamente alejados de este precisamente por la estructura típica de una oración subordinada. No obstan-

te, debido a que dicha estructura se encuentra muy raras veces en los ítems, no fue posible estimar su efecto. Al igual que con las negaciones, se recomienda construir ítems con dicha estructura para poner a prueba su posible efecto en la dificultad.

Una de las hipótesis que más problemas planteó en cuanto a su operacionalización fue la que proponía que el vocabulario de difícil comprensión aumentaría la dificultad de los ítems cuando estuviera presente en la opción correcta y en el texto principal. En la literatura consultada se encontró que quien contesta un ítem necesita conocer el significado de las palabras que lo componen. Cuando dicho conocimiento léxico no existe, la comprensión se empobrece en la medida en que la persona no puede utilizar toda la información requerida para resolver un ítem. Particularmente, cuando uno o varios términos poco conocidos (es decir, de difícil comprensión) están en la respuesta correcta, esta tiende a ser considerada como poco plausible, de modo que los distractores con léxico más accesible se vuelven mejores candidatos (desde la perspectiva de quien contesta) a ser la opción correcta. En la presente investigación, este aspecto se abordó mediante dos estrategias: utilizando dos corpus léxicos como referencia y acudiendo al criterio de jueces expertos.

En cuanto al uso de los corpus léxicos (Sánchez, s.f.), se encontró que cuando uno o más términos están presentes en el ítem y ausentes en los corpus, el nivel de dificultad es menor. Lo anterior no es esperable a la luz de la investigación consultada; por otro lado, se debe tomar en cuenta que en el modelo de regresión estas variables presentaron coeficientes estandarizados por debajo de .10 en dos de los cuatro formularios. Por ello, no es recomendable sacar conclusiones a partir de estos resultados. Ahora bien, como se mencionó anteriormente, la dificultad léxica evaluada por el juez 1 resultó ser sumamente importante y consistente a lo largo de los cuatro formularios.

Dada la magnitud y dirección de los coeficientes estandarizados, es muy importante destacar que el dominio léxico juega un papel crucial en la resolución de estos ítems. Este resultado está acorde con las investigaciones consultadas, en las cuales se plantea que las tareas verbales están sumamente influenciadas por el conocimiento léxico de los sujetos que las resuelven. Por otro lado, una de las características esenciales del léxico es que su comprensión depende de las inferencias que realizan los sujetos, quienes necesitan activar una gran cantidad de conocimiento del mundo para procesarlo (Escudero, 2010).

Antes de dar por concluido este documento, es importante señalar algunas limitaciones de este trabajo que para próximas investigaciones se podrían prever en aras de poner a prueba otro tipo de variables sobre la comprensión de lectura, así como una relación de recomendaciones que desde esta investigación pueden derivarse.

En primer lugar, en esta investigación no se incluyeron variables asociadas a las inferencias y estrategias de resolución que podrían afectar la dificultad de los ítems (Green y Smith, 1987). Esto es importante porque en algunas investigaciones (Tatsuoka 1982 como se citó en Green y Smith, 1987; Gilhooly, 2004) la dificultad de los ítems es vista como una función de las estrategias para resolver el ítem. De esta manera, el poder predictivo de modelos basados en la manera en que los y las estudiantes resuelven los ítems verbales de la PAA sería mayor. En este sentido, de igual forma a como lo plantean Gorin y Embretson (2006, p. 49), “el estudio actual sugiere al menos una lista inicial de variables como potenciales fuentes de dificultad para la construcción de ítems de comprensión de lectura, sin embargo, la cantidad de varianza explicada sugiere que los investigadores deben pensar más allá de los “sospechosos” usuales. Se podría implementar en futuras investigaciones un diseño en el que se observe qué estrategias de resolución utilizan los y las estudiantes cuando encuentran una palabra desconocida o un co-

necto en el texto principal o en las opciones de respuesta.

Es claro que la comprensión de lectura involucra un aspecto inferencial muy importante (Iza y Ezquerro, 2000), el cual deberá ser incorporado en futuras investigaciones para entender mejor qué elementos provocan la variación de los índices de dificultad de los ítems verbales de la PAA. Inclusive, será necesario tomar en cuenta aspectos sociodemográficos de quienes contestan los ítems, ya que podrían encontrarse diferencias entre hombres y mujeres, así como también entre quienes provengan de un colegio público o de uno privado.

Una última limitación que deberá subsanarse en futuras investigaciones es la relacionada con la operacionalización de variables como nivel de abstracción, similitud entre texto relevante para contestar el ítem y respuesta correcta, etc.; por ejemplo, en lugar de nivel de abstracción se podría utilizar la imaginabilidad del texto. Este aspecto es crucial a la hora de trabajar con jueces expertos, ya que una operacionalización inadecuada impide que haya un mínimo de concordancia entre los jueces. Igualmente, será necesario trabajar con consensos inter-jueces, de modo que cada juez trabaje de manera independiente y, al final de su labor, lleguen a consensos.

En cuanto a las recomendaciones que se derivan a partir de esta investigación, se pueden señalar las siguientes medidas para reducir a corto plazo la importancia del componente léxico en la resolución de los ítems verbales de la PAA:

- Revisión de los ítems por parte de población meta y de jueces expertos. Un mecanismo para detectar palabras de difícil comprensión es mostrar los ítems a un grupo pequeño de estudiantes de secundaria y de jueces expertos. Mediante esta retroalimentación, se lograrían identificar aquellos ítems cuya dificultad fuera meramente un asunto de dominio de vocabulario.
- Conformación de una “lista negra” de palabras. Dicha lista se derivaría de las revi-

siones hechas por los jueces y los(as) estudiantes, de modo que las palabras incluidas no se utilizarían en la construcción de nuevos ítems.

- Incorporación de un glosario a la PAA. Otra forma de evitar que el componente léxico interfiera en la medición del constructo es incluir un listado de vocablos con sus respectivas definiciones, de modo que todos los(as) estudiantes tengan el mismo conocimiento de palabras particularmente difíciles de comprender.
- Dado el nivel de precisión alcanzado por el juez uno, podría encargarse de revisar los ítems verbales, a la búsqueda de vocabulario que pueda ser inaccesible para la población de estudiantes.
- Aumentar la extensión del texto principal de los ítems. La gran mayoría de ítems analizados en esta investigación no superaba los tres o cuatro renglones, lo reduce considerablemente la posibilidad de que el o la estudiante obtenga la información necesaria para resolver el ítem. Si los textos tuvieran una mayor extensión, los constructores de ítems podrían incorporar más datos y así la importancia de conocer una palabra clave se reduciría.

## Referencias

- Attoresi, H.F., Picón, J., Abal, F., Aguerri, M. & Galibert, M.S. (2009). Aplicación del modelo LLTM de Fisher al análisis de las fuentes de dificultad de ítems de razonamiento deductivo. *Interdisciplinaria*, 26(1), 77-93.
- Baker, F. (1993). Sensitivity of the Linear Logistic Test Model to Misspecification of the Weight Matrix. *Applied Psychological Measurement*, 17(3), 201-210. doi: [10.1177/014662169301700301](https://doi.org/10.1177/014662169301700301)
- Belinchón, M., Igoa, J. & Rivière, A. (1998). *Psicología del lenguaje: Investigación y teoría*. España: Editorial Trotta.
- Bond, T. & Fox, C. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Estados Unidos: Lawrence Erlbaum Associates, Inc.

- Borsboom, D., Mellenbergh, G.J. & Heerden, J. (2004). The Concept of Validity. *Psychological Review*, 111(4), 1061–1071. doi: [10.1037/0033-295X.111.4.1061](https://doi.org/10.1037/0033-295X.111.4.1061)
- Chen, Y., MacDonald, G. & Leu, Y. (2011). Validating Cognitive Sources of Mathematics Item Difficulty: Application of the LLTM to Fraction Conceptual Items. *The International Journal of Educational and Psychological Assessment*, 7(2), 74-93. Recuperado de <https://sites.google.com/site/tijepa2012/home>
- Clark-Carter, D. (2004). *Quantitative Psychological Research*. Estados Unidos: Psychology Press.
- Cortada de Kohan, N. (2003). Posibilidad de integración de las teorías cognitivas y la psicometría moderna. *Revista Argentina de Neuropsicología*, 1, 8-23.
- Cronbach, L.J. & Meehl, P.E. (1955). Construct Validity in Psychological Test. *Psychological Bulletin*, 52(2), 281-302.
- Daneman, M. & Hannon, B. (2001). Using Working Memory Theory to Investigate the Construct Validity of Multiple-Choice Reading Comprehension Tests Such as the SAT. *Journal of Experimental Psychology: General*, 130(2), 208-233.
- Davey, B. (1988). Factors Affecting the Difficulty of Reading Comprehension Items for Successful and Unsuccessful Readers. *The Journal of Experimental Education*, 56(2), 67-76.
- De Boeck, P. & Wilson, M. (2004). *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. NY: Springer.
- DeMauro, G., Merritt, A. & Adams, R. (1994). Delimiting the Verbal Domain (Research Report RR-94-34). Princeton, NJ: Educational Testing Service.
- Drum, P.A., Calfee, R.C. & Cook, L.K. (1981). The Effects of Surface Structure Variables on Performance in Reading Comprehension. *Reading Research Quarterly*, 16(4), 486-514.
- Embretson, S. & Daniel, R. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. *Psychology Science Quarterly*, 50(3), 328-344.
- Embretson, S. E. & Wetzel, D. (1987). Component Latent Trait Models for Paragraph Comprehension Tests. *Applied Psychological Measurement*, 11(2), 175-193. doi: [10.1177/014662168701100207](https://doi.org/10.1177/014662168701100207)
- Embretson, S. E. (1996). The New Rules of Measurement. *Psychological Assessment*, 8(4), 341-349.
- Embretson, S. E. (2010). Cognitive Design Systems: A Structural Modeling Approach Applied to Developing a Spatial Ability Test. En S. Embretson (Ed.), *Measuring Psychological Constructs: Advances in Model-Based Approaches* (pp. 247-271). Washington, DC: American Psychological Association.
- Embretson, S.E. & Gorin, J. (2001). Improving Construct Validity with Cognitive Psychology Principles. *Journal of Educational Measurement*, 38(4), 343-368. doi: [10.1111/j.1745-3984.2001.tb01131.x](https://doi.org/10.1111/j.1745-3984.2001.tb01131.x)
- Embretson, S.E. (1989). Latent Trait Models as an Information-Processing Approach to Testing. *International Journal of Educational Research*, 13(2), 189-203. doi: [10.1016/0883-0355\(89\)90006-2](https://doi.org/10.1016/0883-0355(89)90006-2)
- Escudero, I. (2010). Las inferencias en la comprensión lectora: una ventana hacia los procesos cognitivos en segundas lenguas. *Revista Nebrija de Lingüística Aplicada*, 7(4), 1-32. Recuperado de <http://www.nebrija.com/revista-linguistica/>
- Farr, R., Pritchard, R. & Smitten, B. (1990). A Description of What Happens When an Examinee Takes a Multiple-Choice Reading Comprehension Test. *Journal of Educational Measurement*, 27(3), 209-226. doi: [10.1111/j.1745-3984.1990.tb00744.x](https://doi.org/10.1111/j.1745-3984.1990.tb00744.x)
- Fischer, G. (1997). Unidimensional Linear Logistic Rasch Model. En W. Van Der Linden y R. Hambleton, *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.
- Fisher, G. H. (1973). The Linear Logistic Test Model as an Instrument in Educational Research. *Acta Psychologica*, 37, 359-374. doi: [10.1016/0001-6918\(73\)90003-6](https://doi.org/10.1016/0001-6918(73)90003-6)
- Freedle, R. & Kostin, I. (1991). The Prediction of SAT Reading Comprehension Item Difficulty for Expository Prose Passages (Research



- Report RR-91-29). Princeton, NJ: Educational Testing Service.
- Freedle, R. & Kostin, I. (1992). The Prediction of GRE Reading Comprehension Item Difficulty for Expository Prose Passages for each of Three Item Types: Main Ideas, Inferences and Explicit Statements (GRE Board Report No. 87-1OP). Princeton, NJ: Educational Testing Service.
- Gilbert, M.S., Picón, J.C., Lozzia, G.S., Aguerri, M.E. & Attorresi, H.F. (2010). Componentes de dificultad de ítems para la evaluación de operaciones lógicas / Una aplicación del modelo LLTM. *SUMMA Psicológica UST*, 7(1), 3-14.
- Gilhooly, K. (2004). Working Memory and Reasoning. En J. Leighton y R. Sternberg (Eds.), *The Nature of Reasoning* (pp. 49-77). Estados Unidos: Cambridge University Press.
- Gorin, J. S. (2007). Reconsidering Issues in Validity Theory. *Educational Researcher*, 36(8), 456-462. doi: [10.3102/0013189X07311607](https://doi.org/10.3102/0013189X07311607)
- Gorin, J.S. & Embretson, S.E. (2006). Item Difficulty Modeling of Paragraph Comprehension Items. *Applied Psychological Measurement*, 30(5), 394-411. doi: [10.1177/0146621606288554](https://doi.org/10.1177/0146621606288554)
- Gorin, J.S. (2005). Manipulating Processing Difficulty of Reading Comprehension Questions: The Feasibility of Verbal Item Generation. *Journal of Educational Measurement*, 42(4), 351-373. doi: [10.1111/j.1745-3984.2005.00020.x](https://doi.org/10.1111/j.1745-3984.2005.00020.x)
- Graesser, A. C., Millis, K. & Zwaan, R.A. (1997). Discourse Comprehension. *Annual Review of Psychology*, 48, 163-189. doi: [10.1146/annurev.psych.48.1.163](https://doi.org/10.1146/annurev.psych.48.1.163)
- Green, K. & Smith, R. (1987). A Comparison of Two Methods of Decomposing Item Difficulties. *Journal of Educational and Behavioral Statistics*, 12(4), 369-381. doi: [10.3102/10769986012004369](https://doi.org/10.3102/10769986012004369)
- Kintsch, W. & van Dijk, T. (1978). Toward a Model of Text Comprehension and Production. *Psychological Review*, 85(5), 363-394. doi: [10.1037/0033-295X.85.5.363](https://doi.org/10.1037/0033-295X.85.5.363)
- Hunt, E. (2011). *Human Intelligence*. Estados Unidos: Cambridge University Press
- Iza, M. & Ezquerro, J. (2000). Elaborative Inferences. *Anales de Psicología*, 16(2), 227-249.
- Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. NY: Cambridge University Press.
- Kirsch, I. & Mosenthal. (1990). Exploring Document Literacy Underlying the Performance of Young Adults. *Reading Research Quarterly*, 25(1), 5-30.
- Leroy, G., Helmreich, S. y Cowie, J.R. (2010). The influence of text characteristics on perceived and actual difficulty of health information. *International Journal of Medical Informatics*. doi: [10.1016/j.ijmedinf.2010.02.002](https://doi.org/10.1016/j.ijmedinf.2010.02.002)
- López, I. & Pita, S. (1999). *Medidas de concordancia: el índice de Kappa*. Recuperado de <http://www.fisterra.com/mbe/investiga/kappa/kappa2.pdf>
- Mair, P. & Hatzinger, R. (2012). *Extended Rasch Modeling: The R Package eRm* [Vignette]. Recuperado de <http://cran.r-project.org/web/packages/eRm/eRm.pdf>
- Martínez, M., Hernández, M. & Hernández, M. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Messick, S. (1995). Validity of Psychological Assessment. *American Psychologist*, 50(9), 741-749. doi: [10.1037/0003-066X.50.9.741](https://doi.org/10.1037/0003-066X.50.9.741)
- Ozuru, Y., Rowe, M., O'Reilly, T. & McNamara, D. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods*, 40(4), 1001-1015. doi: [10.3758/BRM.40.4.1001](https://doi.org/10.3758/BRM.40.4.1001)
- Poinstingl, H. (2009). The Linear Logistic Test Model (LLTM) as the methodological foundation of item generating rules for a new verbal reasoning test. *Psychology Science Quarterly*, 51(2), 123-134.
- Powers, D.E. & Wilson, S. (1995). Answering the New SAT Reading Comprehension Questions without the Passages. *Journal of Educational Measurement*, 32(2), 105-129. doi: [10.1111/j.1745-3984.1995.tb00458.x](https://doi.org/10.1111/j.1745-3984.1995.tb00458.x)
- Real, E., Olea, J., Ponsoda, V., Revuelta, J. & Abad, F. (1999). Análisis de la dificultad de un test de matemáticas mediante un modelo componencial. *Psicológica*, 20, 121-134.
- Revuelta, J. & Ponsoda, V. (1998). Un test adaptativo informatizado de análisis lógico

- basado en la generación automática de ítems. *Psicothema*, 10(3), 709-716.
- Ríos, G. (2011). *Características del lenguaje de los jóvenes costarricenses desde la disponibilidad léxica*. España: Ediciones Universidad de Salamanca.
- Romero, S., Ponsoda, V. & 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.
- Rupp, A., Ferne, T. & Choi, H. (2006). How Assessing Reading Comprehension with Multiple-Choice Questions Shapes the Construct: A Cognitive Processing Perspective. *Language Testing*, 23(4), 441-474. doi: [10.1191/0265532206lt337oa](https://doi.org/10.1191/0265532206lt337oa)
- Sánchez, V. (s.f.). *Corpus de frecuencia de uso de vocabulario en estudiantes de secundaria [datos no procesados]*. Universidad de Costa Rica: Estudios de Lexicografía Hispano-Costarricense.
- Sheehan, K. & Mislevy, R. (1990). Integrating Cognitive and Psychometric Models to Measure Document Literacy. *Journal of Educational Measurement*, 27(3), 255-272. doi: [10.1111/j.1745-3984.1990.tb00747.x](https://doi.org/10.1111/j.1745-3984.1990.tb00747.x)
- Sheehan, K. M., & Ginther, A. (2001). *What do Passage-Based Multiple-Choice Verbal Reasoning Items Really Measure? An Analysis of the Cognitive Skills Underlying Performance on the Current TOEFL Reading Section*. Paper presented at the 2000 Annual Meeting of the National Council of Measurement in Education, New Orleans, LA.
- Sonnleitner, P. (2008). Using the LLTM to Evaluate an Item-Generating System for Reading Comprehension. *Psychology Science Quarterly*, 50(3), 345-362.
- Stahl, S.A. & Jacobson, M.J. (1986). Vocabulary Difficulty, Prior Knowledge, and Text Comprehension. *Journal of Reading Behavior*, 28(4), 309-323. doi: [10.1080/10862968609547578](https://doi.org/10.1080/10862968609547578)
- Tabachnick, B. & Fidell, L. (2007). *Using Multivariate Statistics*. New York: Pearson Education.

## ABOUT THE AUTHORS / SOBRE LOS AUTORES

**Brizuela, Armel** ([armel9@gmail.com](mailto:armel9@gmail.com)). Filólogo y especialista en métodos de investigación cuantitativa en el campo de la psicología. Labora como investigador en el Programa Permanente de la Prueba de Aptitud Académica y en el Programa de Pruebas Específicas para Ingreso a Carrera (adscritos al Instituto de Investigaciones Psicológicas de la Universidad de Costa Rica). [Buscar otros artículos de este autor en Google Académico /](#)

[Find other articles by this author in Scholar Google](#)



**Montero-Rojas, Eiliana** ([eilianamontero@gmail.com](mailto:eilianamontero@gmail.com)). Doctora en Medición y Evaluación Educativa de la Universidad Estatal de Florida, USA, y bachiller en Estadística de la Universidad de Costa Rica (UCR). Catedrática de la UCR desde 1993, se desempeña como docente e investigadora en la Escuela de Estadística, en el Instituto de Investigaciones Psicológicas y en diversos programas de posgrado. Autora o coautora de más de 30 publicaciones científicas. Sus áreas de trabajo incluyen modelos mixtos (multinivel) y modelos de ecuaciones estructurales, evaluación de impacto, modelos de medición, especialmente IRT y Rasch, y validación de instrumentos e indicadores. Su experiencia en consultorías incluye a las siguientes organizaciones: ETS (Educational Testing Service) de Estados Unidos, Fundación Jacobs (Suiza), INCAE Business School, UNAIDS, Universidad de Chile, Escuela Andaluza de Economía Social, Programa Estado de la Educación de Costa Rica, Ministerio de Educación de Costa Rica y Ministerio de Educación de Nicaragua. La American Statistical Association de Estados Unidos le otorgó la distinción "Educational Ambassador 2010-2011". [Buscar otros artículos de esta](#)

[autora en Google Académico / Find other articles by this author in Scholar Google](#)



Brizuela, Armel & Montero-Rojas, Eiliana (2013). Predicción del nivel de dificultad en una prueba estandarizada de comprensión de lectura: aportes desde la psicometría y la psicología cognitiva. *RELIEVE*, 19 (2), art. 1. DOI: [10.7203/relieve.19.2.3143](https://doi.org/10.7203/relieve.19.2.3143)

## ARTICLE RECORD / FICHA DEL ARTÍCULO

<b>Reference / Referencia</b>	Brizuela, Armel & Montero-Rojas, Eiliana (2013). Predicción del nivel de dificultad en una prueba estandarizada de comprensión de lectura: aportes desde la psicometría y la psicología cognitiva. <i>RELIEVE</i> , v. 19 (2), art. 1. DOI: <a href="https://doi.org/10.7203/relieve.19.2.3143">10.7203/relieve.19.2.3143</a>
<b>Title / Título</b>	Predicción del nivel de dificultad en una prueba estandarizada de comprensión de lectura: aportes desde la psicometría y la psicología cognitiva. [ <i>Prediction of the difficulty level in a standardized reading comprehension test: contributions from cognitive psychology and psychometrics</i> ].
<b>Authors / Autores</b>	Brizuela, Armel & Montero-Rojas, Eiliana
<b>Review / Revista</b>	RELIEVE (Revista ELectrónica de Investigación y EValuación Educativa), v. 19, n. 2
<b>ISSN</b>	1134-4032
<b>Publication date / Fecha de publicación</b>	2013 ( <b>Reception Date</b> : 2013 May 30 ; <b>Approval Date</b> : 2013 December 13. <b>Publication Date</b> : 2013 December 16)
<b>Abstract / Resumen</b>	<p><i>This research seeks to identify possible predictors of the difficulty level of reading comprehension items used in a standardized psychometric test for university admission. Several potential predictors of difficulty were proposed, namely, propositional density, negations, grammatical structure, vocabulary difficulty, presence of enhancement elements (words highlighted typographically), item abstraction level and degree of similarity between correct option and relevant text to resolve the item. By Linear Logistic Test Model (Fisher, 1973) it was found that the number of propositions, the syntactic structure, and fundamentally, the presence of difficult words contributed to the prediction of the item difficulty level.</i></p> <p>Esta investigación busca identificar posibles variables predictoras del nivel de dificultad de los ítems de comprensión de lectura utilizados en una prueba psicométrica estandarizada para la admisión a una institución universitaria. Se propusieron varios posibles predictores del nivel de dificultad, a saber: densidad proposicional, negaciones, estructura sintáctica, dificultad del vocabulario, presencia elementos de realce (palabras resaltadas tipográficamente), abstracción del ítem y grado de similitud entre opción correcta y texto relevante para resolver el ítem. Mediante el Modelo Logístico Lineal de Rasgo Latente se encontró que la cantidad de proposiciones, la estructura sintáctica y, fundamentalmente, la presencia de léxico difícil de comprender contribuyeron a la predicción del nivel de dificultad.</p>
<b>Keywords / Descriptores</b>	<p><i>Cognitive psychology, language processing, reading comprehension, Item Response Theory, Linear Logistic Test Model, Task Analysis, Item difficulty level.</i></p> <p>Psicología cognitiva, Procesamiento del lenguaje, Comprensión de lectura, Teoría de Respuesta al Ítem, Modelo Logístico Lineal de Rasgo Latente, Análisis de tareas, Nivel de dificultad de los ítems.</p>
<b>Institution / Institución</b>	Universidad de Costa Rica (Costa Rica).
<b>Publication site / Dirección</b>	<a href="http://www.uv.es/RELIEVE">http://www.uv.es/RELIEVE</a>
<b>Language / Idioma</b>	Español & English version (Title, abstract and keywords in English & Spanish)

## RELIEVE

**Revista ELectrónica de Investigación y EValuación Educativa**  
*E-Journal of Educational Research, Assessment and Evaluation*

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).