

# Evaluación de la Comprensión Lectora mediante un Modelo de Diagnóstico Cognitivo

## Assessment of Reading Comprehension through a Cognitive Diagnosis Model

Pablo Cáceres Serrano<sup>1</sup> y Jesús M. Alvarado Izquierdo<sup>2</sup>

### Resumen

Los Modelos de Diagnóstico Cognitivo (CDM) son modelos confirmatorios de variable latente discreta especialmente indicados para la evaluación de constructos multidimensionales como la comprensión lectora. En el presente trabajo se muestran los pasos para diseñar una evaluación basada en los CDM. En el estudio participaron 434 estudiantes de edades comprendidas entre los 9 y 10 años a los que se aplicaron un conjunto de ítems de comprensión lectora. Una vez determinadas las habilidades que subyacen a los ítems por un panel de expertos, se comparó entre distintos modelos CDM (DINA, DINO, RRUM y G-DINA), mostrando G-DINA un buen ajuste, lo que hizo posible hacer una descripción de los estudiantes en función de sus habilidades en las distintas dimensiones de comprensión lectora y contribuir al conocimiento que se tiene de los atributos a la base de la comprensión.

**Palabras clave:** modelos de diagnóstico cognitivo, clases latentes, comprensión lectora

### Abstract

Cognitive Diagnostic Models (CDM) are discrete latent variable models of confirmatory nature, devised for the evaluation of multidimensional constructs such as reading comprehension. In the present work the steps to design an evaluation of the CDM are shown. The study involved 434 students between the ages of 9 and 10 years old who were given a set of reading comprehension items. A panel of experts identified the skills underlie to the different items. After the identification of the set of skills, it was built a matrix of items by skills, called Q-matrix. Upon base on Q-matrix, it was studied the fit between different CDM models (DINO, DINA, RRUM and G-DINA) obtaining a good fit to the G-DINA model. This result allowed to make a description of the students according to their abilities in the different dimensions of reading comprehension and to contribute to the knowledge of the skills underlie to comprehension.

**Keywords:** cognitive diagnostic models, latent classes, reading comprehension

<sup>1</sup> Psicólogo, Doctor Metodología de la Investigación. Centro Educación Inclusiva, Pontificia Universidad Católica de Valparaíso Académico. Avenida El Bosque 1290, CP 2530388, Viña del Mar, Chile. Tel.: 322274361. Correo: pablo.caceres@pucv.cl

<sup>2</sup> Psicólogo, Doctor en Psicología. Facultad de Psicología, Universidad Complutense de Madrid. Académico Campus de Somosaguas, 28223 – Pozuelo de Alarcón, Madrid, España. Tel.: 913943055. Correo: alvarado@psi.ucm.es

## Introducción

La comprensión de lectura es una competencia clave en el desempeño de las diversas materias escolares y en el desarrollo académico general de los estudiantes en todas las etapas e incluso a nivel universitario (Washer, 2007). Tal es su importancia, que se la incluye en muchas evaluaciones internacionales, como en PISA (Programme for International Student Assessment), PIRLS (Progress in International Reading Literacy Study) (Mullis, Martin, Foy, & Hooper, 2017) o TERCE (Tercer Estudio Regional Comparativo y Explicativo) (UNESCO, 2015). Sin embargo, estas evaluaciones, así como aquellas que se realizan a nivel nacional o local, ofrecen medidas sobre el desempeño de los estudiantes cuyo propósito es posicionar a los sujetos en un continuo o escala, en que las mediciones son entendidas como una suma de puntuaciones (de la Torre y Minchen, 2014). En este tipo de evaluación los ítems miden una única dimensión latente de tipo continuo (Wu, Tam, & Jen, 2016). Para elaborar estas medidas se suele aplicar la Teoría Clásica de los Tests (TCT) o bien alguno de los modelos de Teoría de Respuesta al Ítem (TRI) unidimensionales (Baker & Kim, 2017).

Cuando se requiere información más detallada de los evaluados en cuanto a las posibles deficiencias que explican sus respuestas correctas e incorrectas, una interesante alternativa son los modelos de diagnóstico cognitivo (CDM, por su acrónimo en inglés). Los CDM permiten evaluar el dominio de más de una variable latente, entendidas como habilidades o atributos cognitivos medidos por los ítems que componen el test (de la Torre, 2009; de la Torre, 2011; DiBello, Roussos, & Stout, 2007). Son modelos probabilísticos multidimensionales confirmatorios de clase latente (Rupp & Templin, 2008) que además de una evaluación más detallada del evaluado, también ayudan a proveer una retroalimentación a los estudiantes y permitir mejorar el subsecuente aprendizaje (Kuo, Chen, Yang, & Mok, 2016). Esta perspectiva es lo que se comprende como evaluación centrada en el estudiante, que implica un análisis detallado de los atributos medidos, permitiendo que el examinado conozca los factores que influyen en su

desempeño, lo que a su vez da al profesor la oportunidad de trabajar fortalezas y debilidades específicas, diagnosticar sus dificultades y proponer, eventualmente, medidas oportunas (Boora, Pasiphol, & Tangdhanakanond, 2015; López-Higes, & Rubio-Valdehita, 2014; Romero & Ordoñez, 2014).

Los CDM parten del supuesto de que la población está formada por sujetos con distintas características o habilidades. El objetivo de los CDM es confirmar la estructura teórica en un test en el que se establecen las habilidades que explican el desempeño de los ítems y que permiten diferenciar las distintas clases de sujetos en función de las habilidades que éstos dominan, conformando patrones de dominio de un constructo que se asume multidimensional (Rupp, Templin, & Henson, 2010).

Las principales diferencias entre los CDM con respecto a la TCT y TRI son las suposiciones de variable latente discreta y multidimensionalidad. Estos modelos, incluidos los modelos multicomponentes como el Modelo Logístico Lineal (LLTM, ver Alvarado, Puente, Fernández, & Jiménez, 2015) asumen que la variable latente es continua y unidimensional, al no existir clases de sujetos se hace posible ordenarlos en una escala unidimensional según su mayor o menor nivel en el rasgo (Alvarado-Izquierdo Santisteban-Requena, 2006).

En el presente trabajo se hace una breve revisión de las aplicaciones de los CDM a la lectura y comprensión lectora, explicándose la lógica y características de estos modelos. El principal objetivo es mostrar cómo realizar una evaluación de la comprensión lectora a partir de los CDM, lo que requiere identificar la composición de atributos cognitivos que están a la base de la medida de comprensión lectora, identificando los patrones de dominio cognitivo que prevalecen entre estudiantes chilenos de cuarto año básico, para finalmente realizar los oportunos contrastes e identificar el modelo CDM con mayor poder explicativo.

### Atributos cognitivos en comprensión de lectura

Romero (2010) indica que los CDM tienen su raíz en el trabajo de Lazarsfeld y Henry (1977), relativo al análisis de la estructura latente, mediante el que se buscaba establecer la relación

entre elementos observables y las clases latentes que permiten clasificar dichos elementos. Embretson (1987) propuso incluir la perspectiva cognitiva con el fin de entender con mayor detalle algunas tareas cognitivamente complejas, como la comprensión lectora, teniendo en cuenta dos aspectos, a saber: a) La representación del constructo y b) las redes nomológicas. Lo primero es entendido como el significado de las puntuaciones de los tests, a lo que se accede mediante la comprensión de procesos y estrategias que son utilizados por los evaluados cuando resuelven un ítem; lo segundo es entendido como la utilidad del test para medir las diferencias individuales (Romero, 2010). Con esta aproximación se mejora la calidad de la información que se obtiene con la aplicación del instrumento. Al mejorar la calidad de la información, no solo se puede identificar con mayor precisión los atributos o habilidades cognitivas comprometidas en el desempeño exitoso que deriva de la medición de un constructo, sino que también permite determinar de modo más detallado las fortalezas y debilidades de los evaluados, lo que favorece la interpretación que es posible hacer de los resultados del test (Lee & Sawaki, 2009).

Producto de la diversidad teórica relativa a comprensión de lectura, las aplicaciones en dicho marco de trabajo no son necesariamente coincidentes (Brunning, Schraw, Norby, & Ronning, 2007), aunque todas aportan al entendimiento de los mecanismos que explican su desempeño. Esta misma dificultad enfrentaron quienes decidieron construir las matrices de atributos para explicar el dominio cognitivo sobre comprensión (Buck, Tatsuoka, & Kostin, 1997; Svetina, Gorin, & Tatsuoka, 2011), quienes después de una revisión exhaustiva hallaron que las habilidades a la base su desempeño eran: Reconocer información, interpretar palabras en contexto, sintetizar información dispersa, procesar texto complejo, entender globalmente un pasaje de texto, aplicar conocimiento previo, aplicar estrategias analógicas, identificar vocabulario de poco uso, mantener información en la memoria y rescatarla para hacer inferencias, mantener información en la memoria de trabajo y detectar similitudes semánticas entre respuesta correcta y distractores. Lee y Sawaki (2009)

dispusieron una estructura más simple de atributos: Comprensión de detalles, conexión de ideas e inferencias. Intentos posteriores han identificado otros atributos, como: Vocabulario, sintaxis, extracción de información explícita y comprensión de información implícita (Li, Hunter, & Lei, 2015), los identificados por Ravand (2016): Lectura de detalles, lectura inferencial (inferencia), lectura de idea principal, sintaxis y vocabulario o los propuestos en el trabajo de Hemmati, Baghaei y Bemani (2016): Generación de inferencias, extracción de información explícita, identificación del significado de palabras de acuerdo al contexto, identificación de referencias pronominales y evaluación de opciones de respuesta. En todos estos casos, el paso siguiente a la identificación y selección inicial de las habilidades o atributos que permitieron explicar el desempeño en comprensión, fue la elección del modelo cognitivo más adecuado al constructo y los datos.

### **Modelos de diagnóstico cognitivo en comprensión de lectura**

Li, Hunter y Lei (2015) revisaron los modelos cognitivos que habían sido utilizados exclusivamente en el contexto de la comprensión de lectura. De acuerdo a estos autores, una decisión de importancia corresponde a la elección de un modelo compensatorio o no compensatorio, también llamado conjuntivo. Estos últimos han sido mucho más comunes en comprensión de lectura (Buck, Tatsuoka, & Kostin, 1997; Buck, Van Essen, Tatsuoka, Kostin, Lutz, & Phelps, 1998; Hemmati, Baghaei, & Bemani, 2016; Jang, 2009; Ravand, Barati, & Widhiarso, 2013; Ravand, 2016; von Davier, 2005). Algunos conocidos modelos de tipo no compensatorio son el Rule Space Model (Svetina, Gorin, & Tatsuoka, 2011; Tatsuoka, 2009), el Attributes Hierarchy Model o AHM (Wang & Gierl, 2011), el modelo Deterministic-Input, Noisy-And-Gate, DINA, el modelo Reparameterized Unified Model (RUM) o su sutil derivación, el Fusion Model (Jang, 2009) y el modelo Reduced Reparameterized Unified Model, o RRUM (Jang, 2009; Jang, Dunlop, Wagner, Kim, & Gu, 2013; Li, Hunter, & Lei, 2015), también reconocido como Noncompesatory Reparameterized Unified Model, NC-RUM (Ravand & Robitzsch, 2015). Por su

parte, un conocido modelo compensatorio es el Deterministic-Input, Noisy-OR-Gate, DINO (Liao, Kuo, & Deenang, 2015; Park & Cho, 2017).

Más recientemente, han sido aplicados con buenos resultados algunos modelos de carácter general o saturados. Entre ellos el denominado Generalized Deterministic-Input, Noisy-And-Gate Model, G-DINA (Chen & Chen, 2016; Hemmati, Baghaei, & Bemani, 2016; Ravand, 2016), que agrupa los modelos DINA, DINO, RRUM y el General Diagnostic Model o GDM (von Davier, 2005).

El modelo G-DINA fue propuesto por de la Torre (2011) como una generalización del modelo DINA, que es de tipo no compensatorio y que clasifica a los examinados en dos grupos para cada uno de los ítems: Aquellos que dominaron todos los atributos requeridos por el ítem  $j$  y aquellos que no dominaron al menos uno de los atributos. DINA es restrictivo, pues no dominar un atributo necesario para responder un ítem, es igual a no dominar ninguno de ellos. G-DINA relaja el supuesto de igual probabilidad de éxito para todos quienes no han dominado cualquiera, alguno o todos los atributos requeridos para un ítem (Ravand, 2016).

Para cada uno de los ítems, el modelo G-DINA divide las clases latentes en  $2^{K_j^*}$  grupos latentes, donde  $K_j^* = \sum_{k=1}^{K_j^*} q_{jk}$  representa el número de atributos requeridos para el ítem  $j$ . Se define  $\alpha_{lj}^*$  como el vector de atributos reducido cuyos elementos son los atributos requeridos para el ítem donde  $l = 1, \dots, 2^{K_j^*}$ , que representa el número total de grupos latentes o patrones de atributos latentes requeridos (patrones de dominio). En el modelo G-DINA, la probabilidad de que un examinado perteneciente a la clase latente  $l$ , con un patrón de atributos  $\alpha_{lj}^*$  responda correctamente un ítem  $j$  se denota como  $P(X_j = 1 | \alpha_{lj}^*) = P(\alpha_{lj}^*)$  (García, Olea y de la Torre, 2014). La formulación original del modelo G-DINA basada en  $P(\alpha_{lj}^*)$  puede ser descompuesta en la suma de los efectos debido a la presencia de atributos específicos y sus interacciones, así:

$$P(\alpha_{lj}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'} \alpha_{lk} \alpha_{lk'} \dots + \delta_{j12\dots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}$$

Donde  $\delta_{j0}$  es el intercepto para el ítem  $j$ ,  $\delta_{jk}$  es el efecto principal debido a  $\alpha_k$ ,  $\delta_{jkk'}$  es el efecto de interacción debido a  $\alpha_k$  y  $\alpha_{k'}$  y  $\delta_{j12\dots K_j^*}$  es el efecto de interacción debido a  $\alpha_1 \dots \alpha_{K_j^*}$ . De los parámetros expuestos, se puede señalar que  $\delta_0$  es la probabilidad basal o probabilidad de respuesta correcta cuando ninguno de los atributos está presente,  $\delta_k$  es el cambio en la probabilidad de una respuesta correcta como resultado de dominar un único atributo (por ejemplo  $\alpha_k$ ),  $\delta_{kk'}$  es un efecto de interacción de primer orden y representa el cambio en la probabilidad de una respuesta correcta debido al dominio tanto en  $\alpha_k$  y  $\alpha_{k'}$  que está más allá del impacto aditivo de dominio de ambos atributos, y  $\delta_{12\dots K_j^*}$  representa el cambio en la probabilidad de una respuesta correcta debido al dominio de todos los atributos requeridos más allá del impacto aditivo de los efectos principales y de interacción. El intercepto es siempre no negativo, los efectos principales son usualmente no negativos, pero los efectos de interacción pueden tomar cualquier valor. Cabe notar que los efectos principales son no negativos si  $P(\mathbf{0}_{K_j^*}) \leq P(\alpha_{lj}^*)$  para  $\sum_{k=1}^{K_j^*} \alpha_{ljk}^* = 1$ , donde  $\mathbf{0}_{K_j^*}$  es el vector nulo de largo  $K_j^*$ . Esto implica que tener dominio sobre cualquiera de los atributos cognitivos requeridos para un determinado ítem representa un incremento en la probabilidad individual de obtener un resultado exitoso en el ítem (de la Torre, 2011).

## Método

### Muestra

Corresponde a una muestra no probabilística intencional en el que participaron tanto estudiantes de escuelas públicas como privadas situadas en las Regiones Metropolitana y de Valparaíso de Chile. Los examinados cursaban 4° Básico, con edades comprendidas entre los 9 y los 10 años. Se evaluó a un total de 501 sujetos que se redujo posteriormente a 434 estudiantes, una vez eliminados los protocolos no respondidos, inaceptablemente incompletos o erróneos. De los 434 estudiantes, 256 (59%) eran de sexo femenino y 178 (41%) de sexo masculino. El número de examinados de escuelas públicas fue de 129

(29.7%) y el de escuelas privadas de 305 (70.3%) estudiantes.

### Diseño

El estudio es cuantitativo, ex post facto y de carácter instrumental ya que se enfoca en las propiedades psicométricas del test desde la perspectiva cognitiva.

### Instrumento

Se aplicó la prueba de Comprensión Lectora y Producción de Textos (CP-PT, Medina y Gajardo, 2014) que evalúa comprensión lectora, producción de textos y manejo de la lengua. El instrumento fue desarrollado bajo la lógica de Teoría Clásica de los Tests (TCT). Su construcción se llevó a cabo mediante dos aplicaciones experimentales sobre una muestra inicial de 300 estudiantes, seguida por otra de 1569 sujetos. La primera aplicación abarcó estudiantes de dos regiones del país (que reúnen al menos el 60% de la población chilena) y permitió obtener descriptivos, análisis de dificultad y fiabilidad de sus escalas. La validez se obtuvo mediante criterio de jueces. La segunda aplicación experimental consideró regiones, nivel socioeconómico y cursos, desde los cinco a los 10 años aproximadamente. Con esta aplicación se obtuvo la dificultad, discriminación, fiabilidad y validez definitiva, así como las normas. La discriminación se obtuvo por correlación ítem-test fijando como valor mínimo .4, aunque se hicieron algunas excepciones hasta .3. La fiabilidad se obtuvo mediante test-retest por dimensión, cuyos valores fluctuaron entre .61 a .82, dependiendo del rango de edad a aplicar. La validez definitiva se obtuvo mediante referencia a criterio, comparando el desempeño de esta prueba con las calificaciones escolares finales, con correlaciones que fluctuaron entre .32 a .51 dependiendo del rango de edad. Las normas se restringen solo a rangos centiles (Medina & Gajardo, 2014). Se trata de la única prueba comercial disponible validada y publicada siguiendo protocolos conocidos (TCT) para medir comprensión lectora en el nivel escolar informado (Neira & Castro, 2013).

### Procedimiento

Siguiendo las recomendaciones de Ravand y Robitzsch (2015) el estudio CDM comienza por la

definición de la matriz Q en la que se indican los atributos que requiere cada ítem para poder contestarlo correctamente. A continuación, se procedió a la depuración de la base de datos para la selección de los ítems más discriminativos, para finalmente especificar los modelos y comparar su bondad de ajuste.

### Definición de la matriz Q

Para construir la matriz, en primer lugar, se definieron los atributos a partir de estudios previos que habían utilizado esta metodología en lectura y comprensión lectora (Buck, Tatsuoaka, & Kostin, 1997; Hemmati, Baghaei, & Bemani, 2016; Li, Hunter, & Lei, 2015; Ravand, 2016). En segundo lugar, se instruyó a una experta en comprensión de textos sobre el uso los CDM y se le pidió que elaborase un modelo atribucional que se adecuara al test utilizado, procediendo por retroajuste, mediante un análisis post hoc del contenido del instrumento (Ravand & Robitzsch, 2015). La experta basó su modelo en la concepción teórica de Kintsch, perspectiva centrada en el procesamiento cognitivo de la información (Tijero, 2009). En tercer lugar, una vez que la experta definió los atributos cognitivos con base en la literatura provista, se envió el resultado a tres jueces independientes, todos académicos del área de lingüística, quienes evaluaron la selección de atributos desde la misma concepción teórica y su asignación a los ítems del test. Para ello se usó un protocolo matricial de asignación de definiciones teóricas de las dimensiones y sus atributos a los ítems del instrumento. Una vez obtenidas las evaluaciones de los jueces, los investigadores redactaron un informe donde se expusieron los puntos de acuerdo y desacuerdo y se le volvió a solicitar a la experta inicial que reconsiderara la selección y la asignación de atributos realizada en aquellos casos en los que era necesaria una nueva valoración. Finalmente, se efectuó una última consulta a todos los jueces sobre el resultado de este proceso. Con base en esta revisión final se obtuvo la matriz Q definitiva.

### Especificación de los modelos

Se aplicaron los modelos recomendados en comprensión lectora: G-DINA, DINA, DINO (Chen & Chen, 2016; Liao, Kuo, & Deenang,

2015; Ravand, 2016) y RRUM (Kim, 2011). Estos modelos permiten determinar la naturaleza compensatoria o conjuntiva (Lee & Sawaki, 2009) de relación entre atributos y patrones de respuesta.

### Comparación de los modelos

La comparación se realizó con el fin de establecer cuál era el que mejor bondad de ajuste mostraba y en consecuencia permitía describir mejor los resultados obtenidos. Para juzgar la bondad de ajuste se utilizaron la razón de verosimilitud, el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC) (Hu, Miller, Huggins-Manley, & Chen, 2016), en los que cuanto menor es el valor del índice mejor es la bondad de ajuste (Ravand & Robitzsch, 2015). También se utilizaron índices absolutos, que comparan el ajuste del modelo a los datos reales. Entre ellos, se incluyó MADcor, o la media de las desviaciones absolutas entre las correlaciones observadas y predichas para cada par de ítems  $i$  y  $j$ , esperando valores iguales o menores a 0.05 (DiBello, Roussos, & Stout, 2007; Lei & Li, 2016), el índice  $\max \chi^2$  que de resultar significativo ( $p < .05$ ) indica que se viola la independencia estadística entre pares de ítems y por tanto el modelo no ajusta bien (Hu, Miller, Huggins-Manley, & Chen, 2016), la raíz cuadrada de la media de residuos estandarizados (SRMSR, de su acrónimo en inglés), que se le considera un promedio de correlaciones residuales, fruto de la diferencia entre la correlación muestral y la correlación esperada entre cada par de ítems  $i$  y  $j$  (Maydeu-Olivares & Joe, 2014), siendo valores aceptables aquellos que sean iguales o menores a .05. Por último, se incluyó el índice  $100 * \text{MADRESIDCOV}$  o  $\text{MADRCOV}$ , que compara covarianzas o correlaciones observadas y predichas de pares de ítems. A menor tamaño de efecto, mejor bondad de ajuste (Ravand & Robitzsch, 2015).

Para la comparación de modelos y evaluar la bondad de ajuste se utilizó el entorno de programación R (R Core Team, 2017), en concreto los paquetes Cognitive Diagnosis Modeling, CDM (George, Robitzsch, Kiefer, Groß, & Ünlü, 2016) y G-DINA: The generalized DINA model framework (Ma & de la Torre, 2018).

## Resultados

El primer paso fue la construcción de la matriz  $Q$ . Se definieron seis habilidades presentes en los ítems. Una definición breve de los atributos se muestra en la Tabla 1.

Se tuvo en cuenta la recomendación de identificar entre tres a 10 atributos cognitivos (DiBello, Roussos, & Stout, 2007; Rupp & Templin, 2008).

En la Tabla 2, se dispone la matriz  $Q$  para todos los ítems de la prueba de comprensión lectora. Código de superficie es un atributo básico en casi todos los ítems, mientras que modelo de situación o habilidades lingüísticas tienen un papel menos relevante en el proceso comprensivo.

### Depuración de ítems: análisis de la dificultad y la discriminación

En la Tabla 3 se han dispuesto las medias ( $M$ ) y desviaciones típicas ( $S$ ) de los ítems, en las dos primeras columnas desde la izquierda. Inicialmente se eliminaron los ítems 14a, 14b y 14c ya que presentaban un alto porcentaje de no respuesta, mayor al 20%.

La matriz  $Q$  fue desarrollada asumiendo que solo se podría dar respuesta correcta a los ítems si se dominaban todos los atributos requeridos por cada tarea o ítem de comprensión lectora. Por ello, el modelo de elección para un primer análisis y posterior depuración de ítems fue el modelo DINA.

El modelo DINA es el más parsimonioso, en el que las iniciales DA “Deterministic Input” indican si los examinados en la clase latente  $c$ , han dominado o no todos los atributos medidos para el ítem  $i$ . “And Gate” se refiere a la relación conjuntiva entre los atributos y “Noisy” se refiere al elemento estocástico del modelo DINA que enfatiza que las respuestas de los examinados no son deterministas, sino probabilísticas. Lo estocástico, en este caso está representado por la inclusión de los parámetros distracción ( $s_{ij}$ ) y conjetura ( $g_{ij}$ ), que implica que un examinado que domine todos los atributos requeridos en un ítem, puede no responderlo por distracción o bien, que aún sin tener dominio sobre dichos atributos responda correctamente solo por azar (Ravand, Barati, & Widhiarso, 2013). La discriminación de

Tabla 1. Atributos subyacentes en la comprensión de lectura en el test CL-PT

Parámetro	Atributo	Descripción
$\alpha_1$	Código de superficie	Es un nivel representacional y de procesamiento en el que se mantiene en memoria de modo exacto lo que el texto entrega. Se mantiene la forma (rasgos formales) y la representación lingüística. El tipo de proceso es el básico de decodificación y reconocimiento de palabras. Implica solo reproducción de contenidos y uso de recuerdo. (Riffo, 2016; Tijero, 2009)
$\alpha_2$	Base de texto	Nivel de procesamiento oracional. Los elementos de una oración leída por un usuario son estructurados mediante micro-proposiciones que están conectadas a nivel local, generando un microestructura textual o coherencia local. Permite realizar inferencias básicas que permiten conectar una idea con otra. Es un nivel representacional que permite parafrasear el texto a través del uso de macro-estrategias. (Hemmati, Baghaei, & Bemani, 2016; Riffo, 2016; Tijero, 2009).
$\alpha_3$	Inferencia	Operación que ayuda a agregar la información faltante de aquello que no expresa el nivel de superficie. Las inferencias participan del proceso de comprensión y provienen de conocimiento general, dominios específicos y representaciones basadas en experiencias personales o lecturas previas. Tienen lugar durante la construcción del modelo de situación (Hemmati, Baghaei, & Bemani, 2016; Li, Hunter, & Lei, 2015; Vieiro & Gómez, 2004).
$\alpha_4$	Modelo de situación	Nivel representacional que permite asumir aprendizaje de lo leído. Hace uso del conocimiento previo y el proceso de inferencia. Representa un momento de integración entre el conocimiento explícito y el modelo de mundo que tiene quien comprende (Gernsbacher & Kaschak, 2013; Piacente, 2009; Tijero, 2009).
$\alpha_5$	Instrucciones retóricas	Representan ayudas a la comprensión del texto. Estas ayudas no se refieren a la temática tratada sino al texto en sí y a la relación que se invita a tener con él. Son marcadores en el texto que deben ser interpretados como instrucciones de procesamiento (Sánchez, 1996).
$\alpha_6$	Habilidades lingüísticas	Involucran procesos cognoscitivos y son parte del desarrollo de un individuo, permitiendo la comprensión y expresión de mensajes, elaboración de ideas, reflexión y solución de problemas (Guarneros & Vega, 2014). Son consideradas generales y por tanto más allá de la modalidad escrita.

un ítem se establece como  $\omega_{1j} = 1 - g_j - s_j$ , que se interpreta como la capacidad del ítem para separar adecuadamente a los examinados que dominan los atributos relevantes para dicha tarea respecto de los que no (George y Robitzsch, 2015). El valor de dificultad, por su parte, puede ser reinterpretado como  $\omega_{2j} = [g_j + (1 - s_j)]/2$ , que informa de la facilidad del ítem o probabilidad promedio de responder correctamente un ítem (George & Robitzsch, 2015).

La mayoría de los ítems denota dificultades moderadas a bajas. La dificultad juega un papel en la discriminación, pues ítems muy fáciles o difíciles tienden a discriminar poco y en consecuencia son menos informativos del nivel del rasgo (Ferrando, 1996) en los modelos de variable latente continua. En los modelos de variable latente discreta, sin embargo, pueden ayudar a identificar patrones de respuestas de alguna clase concreta de sujetos.

Los valores de  $g_j$  indican la proporción de examinados que responden correctamente aunque no posean el dominio de todos los atributos relevantes para el ítem en cuestión. Los valores

de  $g_j$  son bastante grandes, en algunos casos llegando a 1, como ocurre con el ítem p13d. En otros casos, como en los ítems p02 o p15c el nivel de conjetura es casi inexistente.

En el caso de  $s_j$ , a pesar de tener todos los atributos relevantes para un determinado ítem, sugiere en general una baja proporción de respuesta incorrecta cuando se tiene el dominio sobre los atributos cognitivos requeridos, con excepciones como en el ítem p02, p10a o p13d.

El valor de discriminación  $\omega_{1j}$ , depende de los dos parámetros anteriores, por lo tanto, si éstos muestran valores bajos, se puede esperar un buen nivel de discriminación. No ocurre con todos los ítems y se aprecian incluso valores negativos. En este último caso, el ítem viola la condición de monotonidad  $g_j < 1 - s_j$ , que asegura que la probabilidad de responder correctamente un ítem, cuando se dominan todos los atributos cognitivos y sin que haya distracción, es mayor que la probabilidad de responder correctamente mediante conjetura aun cuando se carezca de dominio sobre al menos uno de los atributos requeridos (George, Robitzsch, Kiefer, Groß, & Ünlü, 2016). Los ítems con valores negativos deben ser eliminados

Tabla 2. Matriz Q de la prueba de comprensión lectora CL-PT

Ítems	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	Ítems	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$
P01a	1	1	1	0	1	0	P06a	1	0	0	0	1	0
P01b	1	1	1	0	1	0	P06b	1	0	0	0	1	0
P01c	1	1	1	0	1	0	P06c	1	0	0	0	1	0
P01d	1	1	1	0	1	0	P06d	1	0	0	0	1	0
P01e	1	1	1	0	1	0	P08	1	0	0	1	0	1
P01f	1	1	1	0	1	0	P09	1	0	1	0	0	0
P01g	1	1	1	0	1	0	P10a	1	1	1	0	0	0
P01h	1	1	1	0	1	0	P10b	1	1	1	0	0	0
P01i	1	1	1	0	1	0	P10c	1	1	1	0	0	0
P01j	1	1	1	0	1	0	P11	1	1	1	0	0	0
P01k	1	1	1	0	1	0	P12a	0	0	0	0	0	1
P01l	1	1	1	0	1	0	P12b	0	0	0	0	0	1
P01m	1	1	1	0	1	0	P12c	0	0	0	0	0	1
P01n	1	1	1	0	1	0	P13a	1	1	0	0	0	0
P01o	1	1	1	0	1	0	P13b	1	1	0	0	0	0
P01p	1	1	1	0	1	0	P13c	1	0	1	0	0	0
P01q	1	1	1	0	1	0	P13d	1	0	0	0	0	0
P01r	1	1	1	0	1	0	P14a	1	1	0	0	0	0
P02	1	0	0	1	0	1	P14b	1	1	0	0	0	0
P03	1	1	0	0	1	0	P14c	1	1	0	0	0	0
P04	1	1	1	0	0	0	P15a	1	0	1	1	0	0
P05	1	0	1	0	0	0	P15b	1	1	1	0	0	0
							P15c	1	1	1	0	0	0

del análisis, así como aquellos con valores inferiores a .30 (George & Robitzsch, 2015).

El resultado del análisis de la dificultad y discriminación de los ítems a partir del modelo DINA contrasta relativamente con el análisis de la dificultad y la discriminación a partir de los índices clásicos de la TCT ya que si bien la correlación entre la dificultad estimada por DINA y el índice de dificultad de la TCT es de .84, respecto a los índices de discriminación TCT (correlación ítem-test) y DINA es de .59, un valor menor debido a que la correlación ítems test de la TCT asume un modelo unidimensional frente a la naturaleza multidimensional de los CDM. No obstante, la correlación entre los índices de discriminación entre el modelo DINA y G-DINA son más similares con una correlación de .80.

Producto de este primer análisis, para no comprometer la bondad de ajuste se decidió no incluir ítems que no discriminaron adecuadamente de forma que el análisis de las siguientes secciones se ejecuta sobre los 21 ítems que se conservan.

### Validación de la matriz Q

La construcción de la matriz  $Q$  a partir de las opiniones de los expertos es un paso fundamental, pero no está exenta de cierta subjetividad, lo que puede llevar a afectar severamente la estimación de los parámetros del modelo y eventualmente, la correcta clasificación de los examinados. Una manera de enfrentar el problema es validar dicha matriz. Se han propuesto varias alternativas, una de las más prometedoras, asociada al modelo G-DINA, es la de los autores de la Torre y Chiu (2016), que sugieren realizar la validación a partir del índice de discriminación  $\zeta_j^2$  estableciendo la proporción de varianza (PVAF) de un determinado patrón de atributos respecto al máximo valor de discriminación que puede obtenerse cuando todos los atributos han sido correctamente especificados. En el presente estudio el valor promedio PVAF fue de .95 de lo que se concluye que la matriz  $Q$  fue adecuadamente especificada según el criterio propuesto por de la Torre y Chiu (2016). Solo en un ítem, el p13a, se observó un valor bajo de .28,



Tabla 3. Dificultad y discriminación de los ítems

Ítems	M	S	$g_j$	$s_j$	$\omega_{1j}$	$\omega_{2j}$	Ítems	M	S	$g_j$	$s_j$	$\omega_{1j}$	$\omega_{2j}$
p01a	.81	.15	.75	.13	.12	.81	p05	.60	.24	.00	.38	.62	.31
p01b	.66	.22	.41	.17	.42	.62	p06a	.68	.22	.67	.33	.00	.67
p01c	.92	.07	.83	.01	.16	.91	p06b	.92	.07	.96	.10	-.06	.93
p01d	.74	.19	.46	.08	.46	.69	p06c	.70	.21	.75	.33	-.08	.71
p01e	.82	.15	.60	.04	.36	.78	p06d	.67	.22	.70	.35	-.04	.68
p01f	.86	.12	.69	.02	.29	.84	p08	.73	.20	.34	.11	.56	.62
p01g	.91	.08	.81	.02	.17	.89	p09	.45	.25	.45	.54	.00	.46
p01h	.68	.22	.33	.09	.59	.62	p10a	.38	.24	.41	.62	-.03	.39
p01i	.78	.17	.48	.02	.50	.73	p10b	.71	.21	.57	.28	.15	.64
p01j	.95	.05	.93	.03	.04	.95	p10c	.75	.19	.75	.25	.00	.75
p01k	.77	.18	.51	.06	.44	.72	p11	.54	.25	.12	.44	.44	.34
p01l	.92	.07	.81	.01	.18	.90	p12a	.93	.07	.77	.02	.21	.87
p01m	.67	.22	.29	.07	.64	.61	p12b	.68	.22	.22	.18	.60	.52
p01n	.75	.19	.50	.07	.44	.71	p12c	.66	.22	.19	.20	.62	.49
p01o	.78	.17	.56	.07	.37	.75	p13a	.77	.18	.43	.21	.36	.61
p01p	.86	.12	.72	.05	.23	.84	p13b	.87	.11	.72	.12	.16	.80
p01q	.67	.22	.25	.07	.68	.59	p13c	.53	.25	.36	.46	.18	.45
p01r	.66	.22	.24	.07	.69	.59	p13d	.27	.20	1.00	.74	-.74	.63
p02	.31	.21	.05	.58	.37	.24	p15a	.78	.17	.66	.20	.15	.73
p03	.72	.20	.71	.27	.02	.72	p15b	.41	.24	.11	.57	.33	.27
p04	.65	.23	.15	.31	.54	.42	p15c	.67	.22	.07	.28	.65	.39

Tabla 4. Índices de ajuste relativos y absolutos para los CDM

Índices	DINA	DINO	RRUM	G-DINA
Número de parámetros estimados de los ítems	42	42	112	266
Log-likelihood	-4702.521	-4723.926	-4540.060	-4412.947
AIC	9615.042	9657.851	9304.120	9357.894
BIC	10044.867	10087.676	9762.599	10446.784
MADcor	.055	.057	.035	.029
$\max\chi^2$ *	17.898 (.004)	26.852 (.000)	9.791 (.368)	9.086 (.540)
SRMSR	.071	.073	.046	.038
MADRCOV	1.105	1.147	.712	.596

\* En  $\max\chi^2$  los valores-p están entre paréntesis

sin embargo, consultados los expertos, se consideró que la atribución de habilidades era adecuada, por lo que se decidió mantener el ítem.

### Índices de ajuste

El análisis de bondad de ajuste incluyó los índices relativos y absolutos comentados previamente. En la Tabla 4 se muestran los índices para los modelos incluidos.

Los modelos DINA y DINO muestran índices inadecuados de bondad de ajuste, debido a que son los más restrictivos frente a los modelos más flexibles RRUM y G-DINA. RRUM es el que mejor bondad de ajuste muestra en los índices bayesianos, especialmente en BIC que penaliza la

falta de parsimonia, no obstante, si comparamos los modelos G-DINA y RRUM en términos de índices absolutos de ajuste, se observa una mejora en favor de GDINA.

### Patrones de atributos obtenidos de acuerdo al modelo

En la Tabla 5 se muestran los porcentajes de sujetos que presentan alguno de los 16 patrones de atributos. Solo se mantienen los patrones con al menos un 1% de respuestas para el modelo que G-DINA. Dichos patrones están representados por un vector de unos y ceros que indican el dominio o no de cada uno de los atributos que aparecen en

Tabla 5. Patrones de atributos en porcentajes estimados por el modelo G-DINA

N°	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	G-DINA
1	0	0	0	0	0	0	1.45
2	0	0	1	0	1	0	2.63
3	0	1	1	0	0	0	2.11
4	0	1	1	0	1	0	10.56
5	0	1	1	0	1	1	3.8
6	1	0	0	0	0	1	1.83
7	1	0	0	1	1	1	2.01
8	1	0	1	0	0	1	2.47
9	1	0	1	0	1	1	5.76
10	1	1	0	0	0	1	1.35
11	1	1	0	1	1	0	1.49
12	1	1	0	1	1	1	8.69
13	1	1	1	0	0	1	8.64
14	1	1	1	0	1	0	1.03
15	1	1	1	0	1	1	32.39
16	1	1	1	1	1	1	3.97

las columnas de la matriz  $Q$ . Así por ejemplo, un examinado  $i$  que tenga un patrón de atributos igual a  $\alpha_i = \{1,0,0,0,0,1\}$  es alguien que domina los atributos código de superficie y habilidades lingüísticas, pero no los atributos restantes que han sido descritos para la prueba CL-PT.

En la Tabla 5 se aprecia que casi un tercio de las respuestas (32.4%) se corresponden con el patrón  $\alpha_i = \{1,1,1,0,1,1\}$ , es decir, son examinados que dominan todos los atributos excepto el cuarto: modelo de situación. El segundo patrón más frecuente es  $\alpha_i = \{0,1,1,0,1,0\}$  que indica no dominio del código de superficie y de las habilidades lingüísticas, además del desconocimiento del modelo de situación. Son pocos los sujetos que dominan todos los atributos (un 3.97%) y menos aún las respuestas que implican desconocimiento de todos los atributos (apenas un 1.45%).

## Discusión

La evaluación de la bondad de ajuste de los modelos muestra que el modelo G-DINA es el que mejor ajusta a los datos de la prueba de comprensión evaluada, mientras que DINO es el modelo que denota el peor ajuste, lo que revela que un modelo compensatorio no es adecuado. Del mismo modo, un modelo no compensatorio estricto en el que sea preciso dominar todos y cada uno de los atributos, como DINA, tampoco es adecuado para describir las respuestas de los evaluados. En otras palabras, dependiendo de la tarea concreta (el ítem) los evaluados pueden aplicar o no estrategias compensatorias.

En cuanto a los porcentajes por patrón de atributos, G-DINA explora un mayor número de parámetros que el resto de modelos, por su mayor flexibilidad para ajustarse al comportamiento evidenciado de los sujetos, agrupando a los examinados en 16 patrones destacados. Se aprecia que un porcentaje importante denota dominio sobre las habilidades código de superficie y base de texto y que dominar la primera parece incluir la segunda. Del mismo modo, modelo de situación parece estar vinculado a la aparición de las instrucciones retóricas. Por el contrario, un porcentaje relevante de los sujetos no domina el atributo modelo de situación, que está ausente en gran parte de los patrones con mayor frecuencia de sujetos.

Desde la concepción teórica propuesta por Kintsch (ver Tijero, 2009) que consideró la comprensión de textos como un proceso propio del “giro cognitivo”, superando la noción de la comprensión como un conjunto de operaciones lineales realizadas por un lector que se asume pasivo, las evidencias obtenidas en este trabajo concuerdan más con la noción de un proceso comprensivo de carácter complejo e interactivo que requiere que se active una importante cantidad de conocimiento por parte del lector y de la generación de un número relevante de inferencias. La contribución empírica de este trabajo da solidez a esta concepción teórica, aunque teniendo presente que la construcción de la matriz  $Q$  se sostiene en la congruente aplicación de las habilidades definidas dentro de esa dimensión teórica.

Las habilidades específicas definidas, tales

como código de superficie, base de texto, inferencia o instrucciones retóricas tienen sustento en estudios previos. En efecto, en el trabajo de Hemmati, Baghaei y Bemani (2016), en el de Li, Hunter y Lei (2015) o Ravand (2016) hay atributos propuestos que concuerdan con los aquí definidos. Inferencia, vocabulario, (código de superficie), sintaxis o más indirectamente, comprensión de información implícita, guardan relación con la concepción cognitiva de comprensión que se aplicó mediante retroajuste.

Otros atributos como las habilidades lingüísticas están representados y son importantes para alcanzar un desempeño adecuado en la evaluación, pero requieren mejor sustento teórico, pues al implicar procesos cognitivos generales, su identificación y definición varía considerablemente en la literatura (Guarneros & Vega, 2014). Por último, en la medición está poco representada el modelo de situación. Dado que estos atributos emergen por retroajuste, no significa que éste sea menos relevante, sino que el instrumento no lo incluyó como habilidad significativa. Teóricamente hay una relación entre el modelo de situación, el proceso de inferencia y los conocimientos previos, pero entre los patrones rescatados parece haber una necesaria relación con las instrucciones retóricas. Quienes las dominan también dominan el modelo de situación.

Las proyecciones de este trabajo pasan por profundizar la aplicación de estos modelos en poblaciones más diversas y de mayor magnitud, así como estudiar con mayor detalle los factores que pueden llevar a este tipo de comportamiento en las evaluaciones de comprensión. Una limitación del estudio es la posible generalización de los resultados a otras poblaciones, lo que requeriría replicar el estudio en otras muestras y con otros instrumentos de comprensión lectora, así como la necesidad de elaborar instrumentos específicamente diseñados desde la perspectiva de los CDM.

De la aplicación de instrumentos diseñados y/o validados mediante CDM se pueden esperar importantes beneficios como el que las debilidades y fortalezas identificadas en los estudiantes puedan ser abordadas desde una perspectiva esencialmente cognitiva, entendiendo que los atributos definidos en la matriz Q se organizan en función de su complejidad lo que

hace posible orientar el apoyo docente en función de las necesidades que emergen de un diagnóstico con este nivel de detalle.

## Referencias

- Alvarado, J. M., Puente, A., Fernández, M. P., & Jiménez, V. (2015). Análisis de los componentes en la adquisición de la lectura en castellano: Una aplicación del modelo logístico lineal. *Suma Psicológica*, 22(1), 45-52. doi:<https://doi.org/10.1016/j.sumpsi.2015.05.006>
- Baker, F. B. y Kim, S. (2017). *The basics of item response theory using R*. New York, NY: Springer.
- Boora, S., Pasiphol, S., & Tangdhanakanond, K. (2015). Development of cognitive diagnostic testing on basic arithmetic operation. *Procedia - Social and Behavioral Sciences*, 191, 769-772. doi:<https://doi.org/10.1016/j.sbspro.2015.04.533>
- Brunning, R. H., Schraw, G. J., Norby, M. N., & Ronning, R. R. (2007). *Psicología cognitiva y de la instrucción*. Madrid: Pearson Educación
- Buck, G., Tatsuoka, K. K., & Kostin, I. (1997). The subskills of reading: Rule-space analysis of a multiple-choice test of second language reading comprehension. *Language Learning*, 47(3), 423-466.
- Buck, G., Van Essen, T., Tatsuoka, K., Kostin, I., Lutz, D., & Phelps, M (1998). *Development, selection and validation of a set of cognitive and linguistic attributes for the SAT I verbal: Analogy section*. Princeton: Educational Testing Service.
- Chen, H., & Chen, J. (2016). Retrofitting non-cognitive-diagnostic reading assessment under the generalized DINA model framework. *Language Assessment Quarterly*, 13(3), 218-230. doi:<https://doi.org/10.1080/15434303.2016.1210610>
- de la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa*, 20(2), 89-97. doi:<https://doi.org/10.1016/j.pse.2014.11.001>
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1), 115-130. doi:<https://doi.org/10.3102/10769986073094>

- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2), 179-199. doi:<https://doi.org/10.1007/s11336-011-9207-7>
- de la Torre, J., & Chiu, C. (2016) A general method of empirical Q-matrix validation. *Psychometrika*, 81(2), 253-273. doi:<https://doi.org/10.1007/s11336-015-9467-8>
- DiBello, L. V., Roussos, L. A., & Stout, W. (2007) Review of cognitively diagnostic assessment and a summary of psychometric models. En C.R. Rao y S. Sinharay (Eds.) *Handbook of Statistics* (pp. 979-1030). Oxford, UK: Elsevier. doi:[https://doi.org/10.1016/S0169-7161\(06\)26031-0](https://doi.org/10.1016/S0169-7161(06)26031-0).
- Embretson, S. E., & Wetzel, C. D. (1987) Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement*, 11(2), 175-193
- Ferrando, P. J. (1996). Evaluación de la unidimensionalidad de los ítems mediante análisis factorial. *Psicothema*, 8(2), 397-410.
- García, P. E., Olea, J., & De la Torre, J. (2014). Application of cognitive diagnosis models to competency-based situational judgment tests. *Psicothema*, 26(3), 372-377.
- George, A. C., & Robitzsch, A. (2015). Cognitive diagnosis models in R: A didactic. *The Quantitative Methods for Psychology*, 11(3), 189-205. doi:<https://doi.org/10.20982/tqmp.11.3.p189>
- George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software*, 74(2), 1-24. doi:<https://doi.org/10.18637/jss.v074.i02>
- Gernsbacher, M., & Kaschak, M. (2013). Text comprehension. En D. Reisberg (Ed.) *The Oxford Handbook of Cognitive Psychology* (pp. 462-474). Londres: Oxford University Press.
- Guarneros, E., & Vega, L. (2014). Habilidades lingüísticas orales y escritas para la lectura y escritura en niños preescolares. *Avances en Psicología Latinoamericana*, 32(1), 21-35. doi:<https://dx.doi.org/10.12804/apl32.1.2014.02>
- Hemmati, S. J., Baghaei, P., & Bemani, M. (2016). Cognitive diagnostic modeling of L2 reading comprehension ability: Providing feedback on the reading performance of Iranian candidates for the university entrance examination. *International Journal of Language Testing*, 6(2), 92-100.
- Hu, J., Miller, D., Huggins-Manley, A. C., & Chen, Y. (2016). Evaluation of model fit in cognitive diagnosis models. *International Journal of Testing*, 16(2), 119-141. doi:<https://doi.org/10.1080/15305058.2015.1133627>
- Alvarado-Izquierdo, J. M., & Santisteban-Requena, C. (2006). La validez en la medición psicológica. Madrid: Editorial UNED.
- Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for fusion model application to LanguEdge assessment. *Language Testing*, 26(1), 31-73.
- Jang, E. E., Dunlop, M., Wagner, M., Kim, Y., & Gu, Z. (2013). Elementary schools ELL's reading skill profiles using cognitive diagnosis modeling: Roles of length of residence and home language environment. *Language Learning*, 63(3), 400-436. doi:<http://dx.doi.org/10.1111/lang.12016>
- Kim, Y. (2011) Diagnosing EAP writing ability using the reduced reparameterized unified model. *Language Testing*, 28(4), 509-541. doi:<http://dx.doi.org/10.1177/0265532211400860>
- Kuo, B., Chen, C., Yang, C., & Mok, M. M. (2016). Cognitive diagnostic models for tests with multiple-choice and constructed-response items. *Educational Psychology*, 36, 1115-1133. doi:<http://dx.doi.org/10.1080/01443410.2016.1166176>
- Lazarsfeld, P. F., & Henry, N. W. (1977). *Análisis de la estructura latente*. Madrid: Instituto de Estudios Políticos.
- Lee, Y., & Sawaki, Y. (2009). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly*, 6(3), 172-189. doi:<http://dx.doi.org/10.1080/15434300902985108>
- Lei, P., & Li, H. (2016). Performance of fit indices in choosing correct cognitive diagnostic models and Q-matrices. *Applied Psychological Measurement*, 40(6), 405-417. doi:<http://dx.doi.org/10.1177/01466216166647954>
- Li, H., Hunter, V., & Lei, P. (2015). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, 33(3),

- 391-409.  
doi:<https://doi.org/10.1177/0265532215590848>
- Liao, C., Kuo, B., & Deenang, E. (2015). *Exploring the mastery pattern of a grade 4 reading comprehension test using cognitive diagnosis model*. Bangkok: Conference Proceedings of International Conference on Advances in Business y Management (ICABM).
- López-Higes, R., & Rubio-Valdehita, S. (2014). ¿Qué variables determinan el nivel lector de un alumno en el segundo ciclo de Educación Primaria y cuál es su valor diagnóstico? *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 1(37), 31-52.
- Ma, W., & de la Torre, J. (2018). GDINA: The generalized DINA model framework. R package version 2.1. Recuperado de <https://CRAN.Rproject.org/package=GDINA>.
- Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, 49, 305-328.  
doi:<https://doi.org/10.1080/00273171.2014.911075>
- Medina, A., & Guajardo, A. M. (2014) *Pruebas de comprensión lectora y producción de textos CL-PT. Kinder a 4° Básico*. Santiago, CL: Ediciones Universidad Católica
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). PIRLS 2016 International Results in Reading. Recuperado desde Boston College, TIMSS y PIRLS International Study Center website: <http://timssandpirls.bc.edu/pirls2016/international-results/>
- Neira, A., & Castro, G. (2013) Análisis de un instrumento estandarizado para la evaluación de la comprensión lectora a partir de un modelo psicolingüístico. *Estudios Pedagógicos*, XXXIX(2), 231-249.
- Piacente, T. (2009) Especificidad de la evaluación psicológica en los procesos de aprendizaje del lenguaje escrito. Evolución y estado actual. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 2(28), 135-148.
- R Core Team (2017). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Recuperado de <https://www.r-project.org/>.
- Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*, 34(8), 765-781. doi:10.1177/0734282915623053
- Ravand, H., Barati, H., & Widhiarso, W. (2013). Exploring diagnostic capacity of a high stakes reading comprehension test: A pedagogical demonstration. *Iranian Journal of Language Testing*, 3(1), 11-37.
- Ravand, H., & Robitzsch, A. (2015). Cognitive diagnostic modeling using R. *Practical Assessment Research y Evaluation*, 20(11), 1-12.
- Riffo, B. (2016). Representaciones mentales en la comprensión del discurso: Del significante lineal al modelo de situación. *Revista Signos*, 49(1), 205-223.
- Romero, S. J. (2010). *Propiedades y aplicaciones del método de las distancias mínimo-cuadráticas (LSDM) para la validación y análisis de atributos cognitivos* (disertación doctoral no publicada). Universidad autónoma de Madrid, Madrid.
- Romero, S. J., & Ordóñez, X. (2014). Validation of the cognitive structure of an arithmetic test with the Least Squares Distance Model (LSDM). *Universitas Psychologica*, 13(1), 333-346. Disponible en: <http://revistas.javeriana.edu.co/index.php/revPsycho/article/view/3671>
- Rupp, A. A., & Templin, J. L. (2008) Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, 6(4), 219-262. doi:<http://dx.doi.org/10.1080/15366360802490866>
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement. Theory, methods, and applications*. New York, NY: Guilford Press.
- Sánchez, E. (1996). Los textos divulgativos como una conversación encubierta: Análisis de los recursos comunicativos de un texto divulgativo. *Infancia y Aprendizaje*, 75, 85-96.

- Svetina, D., Gorin, J. S., & Tatsuoka, K. K. (2011). Defining and comparing the reading comprehension construct: A cognitive-psychometric modeling approach. *International Journal of Testing, 11*, 1-23.
- Tatsuoka, K. K. (2009). *Cognitive assessment. An Introduction to the rule space method*. New York, NY: Routledge.
- Tijero, T. (2009) Representaciones mentales: Discusión crítica del modelo de situación de Kintsch. *Onomázein, 19*(1), 111-138
- UNESCO (2016). *Informe de resultados TERCE. Tercer estudio regional comparativo y explicativo. Logros de aprendizaje*. Santiago: Organización de Naciones Unidas.
- Vieiro, P. y Gómez, I. (2004). *Psicología de la lectura: Procesos, teorías y aplicaciones instruccionales*. Madrid: Pearson.
- von Davier, M. (2005). *A general diagnostic model applied to language testing data*. Princeton, NJ: Educational Testing Service.
- Wang, C., & Gierl, M. J. (2011). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading. *Journal of Educational Measurement, 48*(2), 165-187.
- Washer, P. (2007). Revisiting key skills: A practical framework for higher education. *Quality in Higher Education, 13*(1), 57-67.
- Wu, M., Tam, H. P., & Jen, T. (2016). *Educational measurement for applied researchers. Theory into practice*. Singapore: Springer.