

---

## Modelo de diagnóstico cognitivo con el entorno R

### *Cognitive Diagnostic Modeling using The R Environment*

Recibido el 3 de octubre de 2018, aceptado el 28 de noviembre de 2018

No. de clasificación JEL: I20; I23; O35

**Álvaro Artavia Medrano**  
Universidad de Costa Rica  
Facultad de Educación  
alvartavia@gmail.com

#### Resumen

Los modelos de diagnóstico cognitivo están basados en estimaciones con variables latentes discretas acerca de los atributos que subyacen a una habilidad global. En contraste con los modelos tradicionales de medición, la evaluación cognitiva se puede interpretar directamente en términos de las competencias que poseen los estudiantes en áreas específicas de conocimiento. Sin embargo, el *software* existente es poco, o bien, se dedica únicamente a un tipo determinado de modelo; asimismo, algunos se han dejado de desarrollar y no se encuentran disponibles, o el costo de la licencia es muy elevado. Se propone, por tanto, la utilización del entorno R, que es un *software* de código abierto y gratuito, en el que se han programado diversos paquetes, tales como CDM, en el que se pueden realizar los principales análisis de una gran variedad de modelos. Se concluye que el potencial de R como entorno para el modelado de diagnóstico cognitivo es un campo promisorio en el que cada vez hay más posibilidades para el desarrollo de algoritmos que permitan interpretaciones más profundas de los datos.

**Palabras clave:** modelado de diagnóstico cognitivo, R, paquete CDM

## **Abstract**

*Cognitive diagnostic models are based on estimates with discrete latent variables about the attributes that underlie a global skill. In contrast to traditional measurement models, cognitive assessment can be interpreted directly in terms of the competencies that students possess in specific areas of knowledge. However, the existing software tools are scarce or dedicated only to a certain type of model. Moreover, some have stopped being further developed and are not available, or the cost of their license is very high. Therefore, the use of the R environment is proposed. This is a free and open source software in which various packages have been programmed. One of them is the CDM package, in which the main analysis of a wide variety of models can be performed. It is concluded that the potential of R as an environment for cognitive diagnostic modeling is a promising field with increasing possibilities for the development of algorithms that allow more extensive interpretations of data.*

**Keywords:** *cognitive diagnostic modeling, R, CDM package*

## **1. Introducción**

Los modelos de diagnóstico cognitivo brindan una realimentación detallada acerca del porqué una persona pudo haber tenido éxito o haber fallado en una prueba determinada. Aunque todavía se les considera novedosos, han estado presentes en la literatura técnica en los últimos 20 años; no obstante, a diferencia de los métodos tradicionales que cuentan con software disponible comercialmente, la evaluación cognitiva tiene muy poco software específico, o bien, el existente se dedica únicamente a un tipo determinado de modelo; asimismo, algunos se han dejado de desarrollar y no se encuentran disponibles, o el costo de la licencia es muy elevado.

En este artículo se presenta una revisión de los modelos de diagnóstico cognitivo, así como se ejemplifica el uso del entorno R y algunas de sus ventajas en los análisis respectivos. Con tal propósito, el documento se ha organizado en cuatro secciones. En la primera de ellas, se exponen los fundamentos teóricos de los modelos de diagnóstico cognitivo, así como algunos ejemplos de sus aplicaciones en diversos ámbitos de la evaluación.

En la segunda, se explicitan los pasos para desarrollar modelos cognitivos. En la tercera, se identifican las ventajas de la utilización del entorno R.

Finalmente, en la cuarta sección se realiza una exposición de cómo trabajar los modelos de diagnóstico cognitivo con la ayuda del paquete CDM y, por último, se presenta una discusión a modo de conclusiones de lo expuesto.

## 2. Referentes

### Modelos de diagnóstico cognitivo

El aprendizaje humano involucra la organización y la reestructuración del conocimiento declarativo y procedimental. Por ello, para diagnosticar problemas en el aprendizaje y brindar información útil para la enseñanza, se requiere identificar los procedimientos cognitivos que facilitan o interfieren en un adecuado desempeño, así como evaluar las estructuras de conocimiento que revelan grados de competencia en un dominio específico de contenidos.

Los atributos, también conocidos como *atributos cognitivos*, son entidades dinámicas que evolucionan a partir del desarrollo de competencias de las personas, por lo que pueden verse como fuentes de complejidad cognitiva en el desempeño de una prueba. De esta manera, “aunque un atributo no es una estrategia, brinda los cimientos para las estrategias. Además, el conjunto de atributos organizados en una estrategia cumple el papel momentáneo de resolver el problema, pero no necesariamente permanecen agrupados como una estrategia” (Leighton, Gierl & Hunka, 2002, p. 4).

Los atributos se pueden definir con distintos grados de especificidad, según el objetivo de la evaluación. Cuanto más fina sea la granularidad con que se quiera hacer afirmaciones acerca de las personas y la identificación de los procesos cognitivos subyacentes, mayor es el alcance de la tarea que se analiza y esto refleja un aumento en la complejidad cognitiva de las acciones involucradas en la resolución (Rupp, Templin & Henson, 2010).

Aunque es posible descomponer atributos complejos en tareas más simples o específicas, esto también aumenta la cantidad de atributos para los ítems, lo cual puede hacer relativamente difícil la estimación de probabilidades de dominio de atributos en un modelo de evaluación cognitiva diagnóstica. Lo recomendable es trabajar una cantidad de atributos que sea estadísticamente manejable de acuerdo con la cantidad de ítems involucrados y el tamaño de la muestra de personas que contestarán la prueba.

Por ejemplo, Roberts, Brito, Gotzmann y Gierl (2009) diseñaron una lista de cinco atributos para elaborar un modelo de aplicación de conocimiento matemático en Álgebra: 1) utilizar letras para representar variables en

expresiones simples, 2) reconocer factores algebraicos, 3) traducir una oración verbal en una expresión algebraica, 4) resolver ecuaciones lineales con una variable, y 5) verificar la solución de una ecuación lineal (por ejemplo, sustituir y simplificar).

Por otra parte, Im y Yin (2009) establecieron siete atributos para diagnosticar competencias, habilidades y conocimientos en Estadística, específicamente en el campo de la prueba de hipótesis. Tales atributos son: 1) establecer hipótesis nulas y alternativas, 2) identificar un método apropiado para el análisis, 3) administrar y llevar a cabo procedimientos complejos de cálculo, 4) realizar procedimientos simples de cálculo, 5) aplicar reglas para pruebas de significancia, 6) utilizar conocimiento y condiciones dadas, y 7) traducir palabras en notación estadística.

Asimismo, Gierl, Leighton, Wang, Zhou, Gokiert y Tan (2009) establecieron una serie de atributos que se requerían para resolver los ítems de álgebra básica I de la prueba SAT del 2005. En particular, tales atributos son: 1) conocimiento matemático básico y habilidades requeridas para establecer una razón simple en la comparación de dos cantidades, 2) dominio de las habilidades para ordenar una serie geométrica, 3) habilidades para resolver series geométricas en un patrón abstracto, 4) habilidades requeridas para representar y ejecutar múltiples destrezas algebraicas, y 5) representación y ejecución de múltiples habilidades avanzadas en álgebra y habilidades de análisis.

Además, Svetina, Gorin y Tatsuoka (2011), con el propósito de desarrollar un modelo cognitivo para la descripción de conocimientos, habilidades y aptitudes en una prueba de lectura crítica, diseñaron una lista de atributos dividida en cinco categorías. Una de ellas es la de habilidades de localización, que permiten a una persona ubicar la información requerida en un texto y comprenderla. Los atributos establecidos para dicha categoría son: 1) utilizar pistas de localización con una línea de referencia, 2) utilizar pistas de localización a partir de una pregunta previa, 3) utilizar pistas de localización que no indican toda la información necesaria, y 4) reconocer información relevante.

También se han desarrollado listas de atributos agrupados según su tipo, por ejemplo, para explicar el rendimiento en la prueba TIMSS-R 99 (Corter & Tatsuoka, 2002; Chen, Gorin, Thompson & Tatsuoka, 2006; Tatsuoka, Guerrero, Corter, Yamada, Tatsuoka, Xin, et al., 2006; Tatsuoka, 2009).

Como se ha indicado, desde un enfoque cognitivo, se asume que la resolución de problemas requiere del procesamiento de la información y de la utilización de secuencias de operaciones o reglas por parte de las personas, de ahí que

se espera que difieran tanto en el conocimiento que poseen como en los procedimientos que utilizan, por lo que se producirá una variabilidad de respuestas en una situación determinada que sea objeto de evaluación (Gierl, Roberts, Brito & Gotzmann, 2009).

Dada la complejidad en la valoración cognitiva del desempeño de una persona en una prueba, se requiere de un modelo en el que sea posible vincular las habilidades evidenciadas en la resolución de problemas con las interpretaciones que se puedan hacer sobre su desempeño. Tal modelo constituye un enfoque con el que se identifican y miden dichas habilidades.

#### Desarrollo de modelos cognitivos

De acuerdo con Gierl, Roberts, Brito y Gotzmann (2009), el término “modelo cognitivo” se refiere a una descripción simplificada de la resolución de problemas en tareas estandarizadas, la cual se hace con algún grado de detalle para facilitar la explicación y la predicción del desempeño de las personas, incluyendo sus fortalezas y debilidades. Los modelos cognitivos son indispensables en la evaluación cognitiva diagnóstica porque brindan un marco de referencia para la interpretación de resultados de tal manera que el desempeño en una prueba se pueda vincular con inferencias específicas acerca del conocimiento y las habilidades de las personas.

Hay muchos beneficios potenciales al modelar el desempeño en una prueba utilizando evaluación cognitiva diagnóstica. La puntuación obtenida en una prueba es un indicador que no permite detallar la forma en que las personas piensan y resuelven situaciones en un área específica de conocimiento. De hecho, frecuentemente se asume que las personas que resuelven correctamente una situación han utilizado el conocimiento y las habilidades apropiados para tal efecto. Sin embargo, este supuesto puede ser falso, ya que se ha demostrado que es posible obtener respuestas correctas utilizando conocimientos y habilidades que no se relacionan con el objetivo especificado en el ítem (Norris, Leighton & Phillips, 2004) y de esta manera las inferencias que se pretendan hacer con respecto a las puntuaciones resultarían inadecuadas, pues se habrían utilizado conocimientos y habilidades que no se pretendían medir en el ítem.

El beneficio de desarrollar una evaluación cognitiva diagnóstica utilizando un modelo cognitivo se refleja en la información detallada que se puede obtener acerca de las estructuras de conocimiento, habilidades de procesamiento y la manera en que son utilizadas por las personas para producir la puntuación de una prueba.

Existen cinco características para un modelo cognitivo de aprendizaje que puede brindar información para el diseño y la interpretación de evaluaciones cognitivas (NRC, 2001):

- 1) estar basado en estudios empíricos acerca de la adquisición de habilidades en el dominio de interés,
- 2) diferenciar el rendimiento entre personas novatas y expertas,
- 3) ser dinámico y lo suficientemente robusto para capturar las diferencias evidenciadas por las personas en cuanto a la comprensión, aprendizaje de conceptos y formas de entender en un dominio específico de contenidos,
- 4) ser lo suficientemente específico como para que los elementos que lo integran apoyen el propósito de la prueba y las inferencias que se pretendan hacer acerca de las puntuaciones.
- 5) permitir un uso flexible tanto para el diseño de pruebas como para la elaboración de informes de resultados.

Por su parte, Leighton y Gierl (2011) ofrecen tres características generales de un modelo cognitivo para la evaluación de los aprendizajes, de tal manera que sea posible vincular las ciencias del aprendizaje con el diseño y la elaboración de pruebas de rendimiento académico. Tales características son:

- 1) Granularidad: el modelo debe magnificar los procesos cognitivos subyacentes al desempeño en una prueba, con el fin de reflejar los tipos de inferencias diagnósticas que se pueden formular a partir del informe de evaluación.
- 2) Mensurabilidad: cada habilidad debe describirse de tal manera que sea posible elaborar ítems que midan dichas habilidades.
- 3) Relevancia para la enseñanza: las habilidades deben ser relevantes y significativas para un amplio grupo de estudiantes y demás personas involucradas en la evaluación; de esta manera, las habilidades que se diagnostiquen sirven como una guía para el mejoramiento de la enseñanza.

Un modelo cognitivo para evaluaciones cognitivas diagnósticas debe reflejar las interrelaciones existentes entre procesos, competencias y habilidades propias de un dominio específico, con el propósito de brindar descripciones basadas en la complejidad creciente de habilidades cognitivas específicas y, de esta manera, relacionar el desempeño de las personas en los ítems de una prueba con sus fortalezas y debilidades.

De acuerdo con Roberts, Brito, Gotzmann y Gierl (2009), tanto el desarrollo como la validación de un modelo cognitivo para evaluación diagnóstica conlleva tres etapas, las cuales se detallan a continuación.

La primera etapa consiste en la elaboración de un modelo inicial para cada una de las áreas de contenido de la asignatura objeto de medición en la prueba seleccionada. Para ello, se utilizan los documentos curriculares que sirvieron de base para elaborar el test. Las habilidades identificadas deben ser de una granularidad fina para poder luego formular inferencias sobre el desempeño de las personas en términos del conocimiento y las habilidades utilizadas.

La segunda etapa permite evaluar los modelos cognitivos desarrollados en la fase previa, para lo que especialistas en contenido pueden revisar los ítems de la prueba y así modificar los modelos preliminares, si se detectan conocimientos y habilidades adicionales o diferentes en dicha revisión.

Esta etapa involucra la evaluación del ordenamiento de las habilidades según su complejidad en términos cognitivos. La revisión utiliza el consenso en los juicios valorativos hechos por especialistas en contenido.

Finalmente, la tercera etapa consiste en utilizar la información de los ítems de la prueba que sirve como base para la elaboración del modelo cognitivo. La idea básica es que especialistas en contenido hagan corresponder los ítems con las habilidades identificadas, lo cual se convierte en una oportunidad para la revisión del modelo en términos de las características deseables en cuanto a su formulación se refiere.

A partir de los planteamientos de Messick (1998), la validez se entiende como un juicio evaluativo integral del grado en que las evidencias empíricas y los supuestos teóricos respaldan qué tan adecuadas y apropiadas resultan ser las inferencias y acciones basadas en modelos de evaluación y en las puntuaciones obtenidas en una prueba.

Al respecto, Gierl (2007) plantea la congruencia de los enfoques de análisis de resultados a partir de atributos cognitivos con los planteamientos de Messick sobre validez, en lo que se considera una forma más accesible y más poderosa de continuar con el desarrollo de la psicología cognitiva y su influencia en la elaboración de pruebas con propósitos diagnósticos. En otras palabras, el tener conciencia de la existencia de diferentes modelos cognitivos puede facilitar la evaluación de las mediciones hechas en campos específicos de conocimiento, con el propósito de generar inferencias diagnósticas, especialmente en términos de procesos de pensamiento, incluyendo errores comunes, fortalezas y habilidades.

Comúnmente, la literatura técnica en torno a los modelos cognitivos para pruebas estandarizadas ha reflejado prácticas que se basan en análisis post hoc, también llamado ajuste posterior [*retrofitting*]; esto es, las pruebas estudiadas no se han elaborado con un propósito cognitivo.

No obstante, Roussos, DiBello, Henson, Jang y Templin (2010) afirman que este tipo de análisis se realizan

(...) usualmente como una demostración de un nuevo modelo estadístico o método o como un intento por extraer mayor información de la que originalmente se podría obtener con el diseño original de la evaluación. En tales casos, el diagnóstico de habilidades esencialmente se convierte en un nuevo propósito adicional para el instrumento de evaluación (p. 38).

Los elementos básicos de un modelo de diagnóstico cognitivo parten de dos elementos: un conjunto de datos con las respuestas a los ítems de una prueba y una matriz, conocida como matriz Q (Tatsuoka, 2009) elaborada por expertos en un área específica de conocimiento.

El conjunto de respuestas es una matriz X de orden  $I \times J$ , donde el elemento  $x_{ij}$  (el elemento de la fila i y la columna j) indica si la persona acertó el ítem ( $x_{ij} = 1$ ) o no ( $x_{ij} = 0$ ).

La matriz Q involucra un trabajo de especialistas que se basan en el área por medir en la prueba para determinar los atributos cognitivos que subyacen a los ítems. Esto se representa en una matriz binaria  $J \times K$ , donde el elemento  $q_{jk}$  expresa si la habilidad se requiere ( $q_{jk} = 1$ ) o no ( $q_{jk} = 0$ ).

Los modelos de diagnóstico cognitivo promueven la inferencia del dominio de atributos por parte de las personas examinadas en una prueba. Si se han determinado K atributos, existirán  $2^K$  combinaciones posibles entre ellos y se conocen como clases de atributos o perfiles de dominio de atributos.

De esta manera, los modelos de diagnóstico cognitivo se basan en tres preguntas fundamentales: 1) ¿cuántas personas tienen un atributo en particular?, 2) ¿cuántas personas tienen una combinación específica de atributos?, y 3) ¿cuáles son los atributos que posee una persona?

Se requiere, por tanto, escoger un modelo específico que brinde las reglas con las que, a partir de los insumos iniciales, se obtengan los resultados para analizar. Asimismo, este modelo determina, entre otros, las probabilidades, las técnicas y los algoritmos para la estimación de los parámetros correspondientes. Si se va a diseñar una prueba con base en un modelo, la elección debe hacerse en el momento de desarrollar los ítems y definir la matriz Q.



Aunque existen varios modelos para llevar a cabo estos análisis, uno de ellos es el modelo DINA [*deterministic input noisy and gate model*] propuesto por de la Torre y Douglas (2004). Este modelo tiene dos propiedades fundamentales: 1) es no compensatorio, es decir, una persona no puede compensar la carencia en un atributo con el dominio de otro, y 2) la probabilidad de acertar un ítem aumenta si la persona domina todos los atributos que se requieren para dicho ítem.

Una persona puede responder incorrectamente un ítem a pesar de que se espere que lo acierte dado que domina los atributos relacionados con el ítem. A esto se le conoce como “equivocaciones” [*slip*]. Por otra parte, puede ser que una persona acierte el ítem por azar, esto es, “adivinando” la respuesta correcta, aunque se espere que no responda correctamente por no dominar los atributos vinculados al ítem. A esto se le llama “adivinación” [*guess*]. El modelo DINA se basa en estos componentes probabilísticos de error, al calcular los parámetros  $g_j$  (adivinación para el ítem  $j$ ) y  $s_j$  (equivocación para el ítem  $j$ ).

Los parámetros  $g_j$  y  $s_j$  reflejan la proporción de personas que muestran respuestas diferentes de las esperadas. Por ejemplo, valores altos de  $g_j$  indican la proporción de respuestas correctas, aunque no se esperara que se dominara el ítem  $j$ ; valores altos de  $s_j$  indican la proporción de respuestas incorrectas para el ítem  $j$ , aunque las personas dominen los atributos requeridos para dicho ítem.

Por otra parte, el valor  $\omega_{1j} = 1 - g_j - s_j$  se conoce como el parámetro de discriminación del ítem y su interpretación es similar a la que se hace en la teoría de respuesta a los ítems, esto es, valores cercanos o mayores que 1 indican una buena separación entre personas con bajas habilidades con respecto a aquellas con altas habilidades.

A medida que  $g_j$  disminuye o  $s_j$  aumenta, la probabilidad promedio de acertar un ítem decrece. El valor  $\omega_{2j} = [g_j + 1 - s_j]/2$  da la idea de la facilidad de un ítem, por lo que corresponde al valor  $p$  del ítem  $j$ , es decir, la proporción de personas que lo responden correctamente.

Dado que los perfiles individuales de atributos no se conocen previamente, los parámetros de los ítems y los perfiles de atributos se deben calcular. Una opción es emplear el método de estimación de probabilidad máxima conjunta [*joint maximum likelihood estimation*], sin embargo, esto puede generar inconsistencias, por lo que el modelo DINA maximiza la probabilidad marginal [*marginal likelihood*] con respecto a los parámetros del ítem y las

probabilidades de dominio de atributos, con lo que se forma un espacio reducido de habilidades.

R como entorno para el análisis de modelos de diagnóstico cognitivo

Una de las limitaciones más grandes para la realización de investigaciones asociadas con modelos de diagnóstico cognitivo ha sido la carencia de software para llevar a cabo los análisis, o no contar con un manual técnico o guía para su utilización, así como contar con una interfaz poco comprensible o que requiere conocimientos técnicos en Linux, entre otros.

Artavia-Medrano (2014), tomó la decisión metodológica de trabajar con una muestra, sobre todo porque los programas de cómputo disponibles en ese momento para este tipo de análisis tienen algunas limitaciones en la cantidad de cálculos que ejecutan. El autor elaboró un modelo con 18 atributos, por lo que tuvo que trabajar con  $2^{18}$  probabilidades de combinaciones de atributos. Dado que contó con 384 sujetos, los cálculos por realizar se basaron en una matriz de orden  $384 \times 2^{18}$ , es decir, 384 filas y  $2^{18}$  columnas, lo cual es un número suficientemente grande para ejecutar los algoritmos programados.

Dicho autor agrega que existen licencias disponibles de manera muy restrictiva y que existen algoritmos propios de un método que son prácticamente indescifrables, dado que no se han dado a conocer públicamente. Para subsanar el faltante de *software* para trabajar con modelos de evaluación cognitiva, comúnmente se han empleado varias aplicaciones a la vez (SPSS, Excel, SAS, Mplus, Mathematica, por ejemplo), muchas de las cuales tienen un costo económico excesivo o demandan grandes conocimientos técnicos por parte de quienes desean investigar en esta área.

Por su parte, Ravand y Robitzsch (2015) advierten que existe software comercial, pero que resulta ineficiente para calcular parámetros de los modelos de diagnóstico cognitivo. Al respecto anotan

Cada ejecución en el software Arpeggio con su número predeterminado de cadenas de Markov y cuatro atributos, por ejemplo, tomaría unos 28 minutos (para 1500 personas y 1000 iteraciones) en una computadora con 2 GB de memoria RAM y un CPU Core i3. Dependiendo de cuántas veces un investigador revise la matriz Q, pasaría horas estimando los parámetros del modelo con el software. Como se mencionó anteriormente, aunque la generación de sintaxis se hace con una macro de SAS, la estimación de parámetros de LCDM se realiza con Mplus, lo que llevaría varias horas con cuatro o cinco atributos (p. 5).

En respuesta a tales limitaciones, se han desarrollado códigos y luego se han conformado paquetes para trabajar con el entorno R. En particular, R es un conjunto integrado de herramientas de software para la manipulación de datos, el cálculo y la visualización gráfica, gratuito y de código abierto. Se caracteriza por ser un lenguaje de programación bien desarrollado, simple y efectivo que incluye condicionales, bucles, funciones recursivas definidas por el usuario y herramientas para el ingreso de datos y sus resultados. De acuerdo con R Core Team (2018),

Muchos usuarios piensan en R como un sistema estadístico. Preferimos pensar en un entorno en el que se implementan técnicas estadísticas. R se puede extender (fácilmente) a través de paquetes. Hay aproximadamente ocho paquetes suministrados con la distribución R y muchos más están disponibles a través de la familia de sitios de Internet CRAN que cubren una amplia gama de estadísticas modernas (párr. 8).

R se puede descargar en <https://cran.r-project.org>, según el sistema operativo que se emplee, así como cuenta con RStudio, disponible en <https://www.rstudio.com>, el cual es un entorno de desarrollo integrado para R, por lo que apoya la ejecución de sus códigos.

La utilización de R en las investigaciones de modelos de diagnóstico cognitivo resulta muy apropiada, básicamente por cuatro razones: 1) es muy eficiente con el tiempo, ya que la estimación de parámetros puede durar menos de un minuto, 2) puede ejecutar gran cantidad de análisis propios de los principales modelos de diagnóstico cognitivo, 3) es gratuito, y 4) los análisis se pueden llevar a cabo con unas cuantas líneas de sintaxis.

Robitzsch, Kiefer, George y Uenlue (2018) desarrollaron el paquete CDM, el cual contiene diversas funciones para el diagnóstico cognitivo, así como modelos multidimensionales de respuesta para ítems dicotómicos y politómicos. En el momento de realizar este artículo, se trabajó con la versión 7.0-12.

En la tabla 1 se exponen algunas propiedades del paquete CDM, las cuales fueron estudiadas con simulación de datos por Rupp y van Rijn (2018).

Tabla 1: Algunas propiedades del paquete CDM

Propiedades de los resultados	
Información de clases latentes	Probabilidades de pertenencia a una clase, consistencia en la clasificación
Información de atributos	Probabilidades de dominio, clasificación de atributos, índice de fiabilidad, correlaciones tetracóricas y policóricas entre atributos

Parámetros de ítems	Adivinación, equivocaciones, índice de discriminación, matrices de índice de discriminación
Salidas gráficas	Probabilidades observadas vs esperadas de atributos, distribución de clases latentes, probabilidad de dominio de atributos para patrones de respuestas y de atributos
Propiedades técnicas adicionales	
Otras	Está basado en un sistema S3, incluye rutinas para la identificación de equivalencia entre clases, identificación de modelos DINA equivalentes, clasificación GDD, creación de ítems pseudodicotómicos a partir de ítems politómicos, creación de espacios reducidos de habilidades, patrones ideales de respuestas, procedimientos Jackknife.

Fuente: André A. Rupp & Peter W. van Rijn (2018) GDINA and CDM Packages in R, p. 73.

### Un ejemplo de utilización del paquete CDM

Estando en R o RStudio, lo primero que se debe hacer es instalar el paquete CDM, para lo que empleamos la función `install.packages("CDM")`; esto se debe hacer una sola vez. No obstante, en cada sesión de trabajo que se vaya a iniciar, se debe cargar el paquete, con la instrucción `library(CDM)`.

Se empleará un conjunto de datos simulados que corresponden a una prueba de 10 ítems y 3 atributos, para 1000 personas simuladas. Tal conjunto se llama `sim10GDINA` y lo contiene el paquete GDINA (Ma, de la Torre, 2018). La matriz Q correspondiente se presenta en la tabla 2.

**Tabla 2: Matriz Q para el conjunto de datos simulados**

Ítem	Atributo 1	Atributo 2	Atributo 3
1	1	0	0
2	0	1	0
3	0	0	1
4	1	0	1
5	0	1	1
6	1	1	0
7	1	0	1
8	1	1	0
9	0	1	1
10	1	1	1

Fuente: Elaboración propia a partir del conjunto de datos simulados

Dado que el paquete CDM contiene el conjunto de datos requerido, así como la matriz Q, lo primero será crear un objeto llamado "misdatos" para realizar las estimaciones del modelo DINA, con la instrucción:

```
misdatos <- din(sim10GDINA$simdat,sim10GDINA$simQ)
```

La función `din()` usa de modo predeterminado el modelo DINA y requiere al menos dos argumentos: el conjunto de datos y la matriz Q. En el caso del conjunto de datos que se está empleando, `simdat` contiene las respuestas simuladas de 1000 examinados y `simQ` es una matriz Q simulada. La convergencia de las estimaciones se produjo con 37 iteraciones, las cuales fueron calculadas en 0.05423808 segundos. Basta con digitar `misdatos` en la consola de RStudio, para tener la información de la cantidad de casos por analizar, el número de ítems, la cantidad de atributos y los patrones de atributos.

La función `IRT.se()` crea un marco de datos con coeficientes, errores estándar e intervalos de confianza del 95% para todos los parámetros del modelo. Con la opción `extended = TRUE`, se solicita el cálculo de la matriz de covarianzas de los parámetros estimados y de los parámetros derivados. Con esto, se crea el objeto “`param`”:

```
param <- IRT.se(misdatos,extended = TRUE)
```

Con el propósito de agrupar los valores por tipo de parámetro, se crea un objeto llamado “`p`” con la instrucción: `p <- split(param,param$partype)`

donde `partype` es un elemento que produjo la función `IRT.se()` y contiene los valores de “adivinación” [`guess`] y “equivocaciones” [`slip`].

Con la instrucción: `pvalues <- colMeans(sim10GDINA$simdat, na.rm=TRUE)`

se determina el porcentaje de personas que acertaron cada ítem de la prueba; en la tabla 2 se muestra dicha información para cada ítem. En la versión 7.0-12 del paquete CDM, `pvalues` es una instrucción, no obstante, se expone su sintaxis con el propósito de poder adecuarla a otros conjuntos de datos. El argumento `na.rm = TRUE` se emplea para obviar los datos faltantes en un conjunto dado.

**Tabla 3: Proporción de personas que acertaron cada ítem**

<u>Ítem</u>	<u>Proporción</u>
1	.55
2	.46
3	.52
4	.46
5	.28
6	.73
7	.43
8	.38
9	.43
10	.45

Fuente: Elaboración propia con el paquete CDM

De acuerdo con lo expuesto en la tabla 3, el ítem 5 resultó el más difícil, dado que fue acertado únicamente por el 28% de las personas examinadas. De igual modo, el ítem 6 se puede considerar como fácil, puesto que el 73% de la población lo contestó correctamente. En este sentido, la dificultad se cataloga de acuerdo con la proporción de examinados que acierta el ítem, es decir, cuanto mayor sea la proporción que lo acierta, menor será la dificultad del ítem y viceversa. Con funciones de los paquetes base de R tales como `mean()` y `sd()` se puede obtener la media aritmética y la desviación estándar, respectivamente.

En el caso que se ejemplifica, se utiliza `mean(pvalues)` y `sd(pvalues)` para obtener `.47` y `.12`, respectivamente; es decir, en promedio, el 47% de las personas examinadas acertó los ítems de la prueba, por lo que se puede concluir que la prueba resultó de dificultad intermedia para la población.

Como se indicó anteriormente, el objeto “p” contiene los parámetros de “adivinación” [*guess*] y “equivocaciones” [*slip*]. Con la instrucción `p$guess$est` se consigue la estimación para el parámetro de adivinación para todos los ítems. Por otra parte, con `p$slip$est` se calcula la estimación del parámetro de equivocaciones para todos los ítems. Al ejecutar ambas instrucciones, se tiene que los parámetros de adivinación están en un rango de `.06` a `.59`, con una media aritmética de `.21` y una desviación estándar de `.16`. Asimismo, los parámetros de equivocaciones están en un rango de `.07` a `.32`, con una media aritmética de `.22` y una desviación estándar de `.08`.

Con la instrucción:

```
omega1 <- 1 - p$guess$est - p$slip$est
```

se calcula la discriminación de cada ítem, para la cual valores cercanos o mayores que 1 indican una buena separación entre personas con bajas habilidades con respecto a aquellas con altas habilidades, en forma similar a la interpretación que se hace en la teoría de respuesta a los ítems. La información producida por esa instrucción permite afirmar que el ítem 3 es el que mejor discrimina (0.82) y que, en términos generales, todos los ítems discriminan bien, aunque el ítem 6 es el que tiene la menor discriminación de todos (0.34).

Por otra parte, con la instrucción:

```
omega2 <- (p$guess$est + (1 - p$slip$est))/2
```

se calcula la facilidad de los ítems, que corresponde al porcentaje de personas que aciertan el ítem y que se calculó con `pvalues`. Por la definición de este parámetro, se pueden encontrar ligeras diferencias con respecto a los valores obtenidos previamente, debidas al redondeo empleado en cada caso.

Ahora bien, para determinar la proporción de personas que poseen un atributo específico, se utiliza la instrucción `p$margprobs`. Sean A1, A2 y A3 los tres atributos de la prueba y P la proporción de personas que poseen cada atributo. Con la instrucción mencionada, se obtiene:  $P(A1) = .71$ ,  $P(A2) = .58$  y  $P(A3) = .54$ ; es decir, el 71% de las personas poseen el atributo 1, el 58% tiene el atributo 2 y el 54% posee el atributo 3. Esto mismo se puede conseguir con la instrucción `misdatos$skill.patt` (solo que en este caso no se tendrán los errores estándar).

Como se sabe, los *ítems* de una prueba no solo miden un atributo, sino una combinación de estos, por lo que también se debe determinar la proporción de personas que poseen una combinación específica de atributos. En este sentido, habrá  $2^K$  combinaciones posibles, donde K es la cantidad de atributos; a esto se le llama distribución de clases de atributos y se determina con la instrucción `p$probs`.

De un modo más directo, esto mismo se puede conseguir con la instrucción `misdatos$attribute.patt`, con la ventaja de que dará la combinación de atributos, la probabilidad de poseer tal combinación y la cantidad de personas examinadas que corresponden a esto. En la tabla 4 se muestra la información.

Tabla 4: Distribución de clases de atributos, errores estándar y frecuencia de examinados

<u>Patrón de atributos</u>	<u>Proporción</u>	<u>Error estándar</u>	<u>Frecuencia de examinados</u>
0 0 0	.08	.02	77
1 0 0	.11	.02	109
0 1 0	.07	.01	70
0 0 1	.06	.01	57
1 1 0	.20	.02	200
1 0 1	.18	.02	176
0 1 1	.08	.01	84
1 1 1	.23	.03	227

Fuente: Elaboración propia con el paquete CDM

Los modelos de diagnóstico cognitivo también permiten obtener el perfil de dominio de atributos para cada persona. Con la instrucción `misdatos$subj.pattern` se obtiene el patrón de respuestas de cada persona examinada; además, con la instrucción `IRT.factor.scores(misdatos, type = "MLE")[1:6,]` se obtiene el perfil de dominio de atributos para las personas examinadas de la 1 a la 6 en la lista. En la tabla 5 se muestra una combinación de la información que se produce con las dos instrucciones explicadas.

Tabla 5: Patrón de respuestas y perfil de dominio de atributos para 6 examinados

ID	Patrón de respuestas	Perfil de atributos
1	1011011101	101
2	1111111101	111
3	1110111010	111
4	1011111110	111
5	0011001000	101
6	1000000101	100

Fuente: Elaboración propia con el paquete CDM

Como se puede notar en la tabla 5, personas con distintos patrones de respuestas pueden tener el mismo perfil de atributos, puesto que las probabilidades de dominio se basan no solo en las respuestas correctas o incorrectas, sino en la dependencia entre *ítems* y atributos que se detalla en la matriz Q. Para saber si los atributos están correlacionados entre sí, se emplea la instrucción `skill.cor(misdatos)$cor.skills`. Los resultados se muestran en la tabla 6.

Tabla 6: Correlación entre los tres atributos según el modelo DINA

	A1	A2	A3
A1	1		
A2	.10	1	
A3	.12	-.04	1

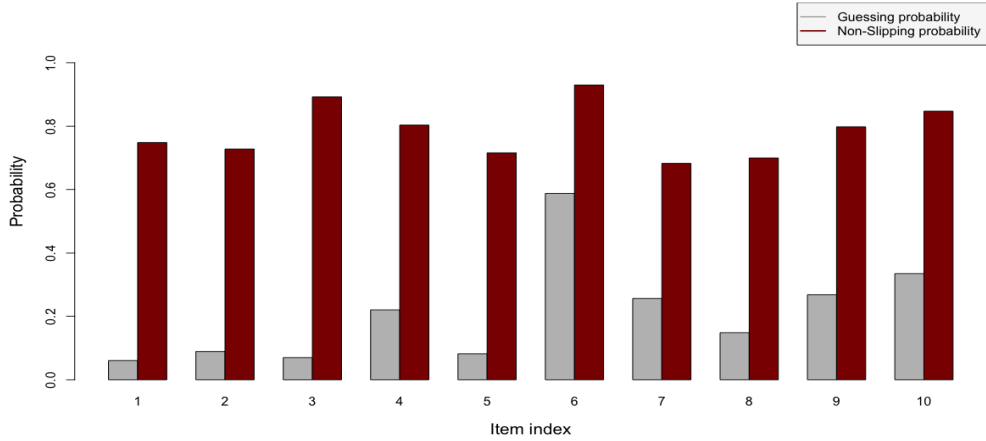
Fuente: Elaboración propia con el paquete CDM

De la tabla 6 se concluye que los atributos no están correlacionados entre sí, por lo que no se puede afirmar que al dominar uno de ellos se tienda a dominar otro. Incluso resulta conveniente observar que existe una correlación negativa entre los atributos 2 y 3. Como son datos simulados y de carácter ilustrativo, únicamente se menciona este hecho, sin embargo, en la interpretación de los datos de una prueba en contextos reales de aplicación, se debe observar con todo cuidado esta situación, puesto que dominar un atributo no debe ir en detrimento del dominio de otro que se está examinando en la misma prueba.

El paquete CDM también ofrece representaciones gráficas. La instrucción `plot(misdatos, pattern = sim10GDINA$simdat[1,])` ofrece las cuatro opciones de gráficos para la persona examinada 1 de la lista. El primer gráfico que se genera representa la estimación de la precisión de los parámetros para cada ítem de la prueba. Las estimaciones se dividen en adivinación y equivocaciones para cada ítem y, en el gráfico 1, se muestra un ejemplo de esta situación.



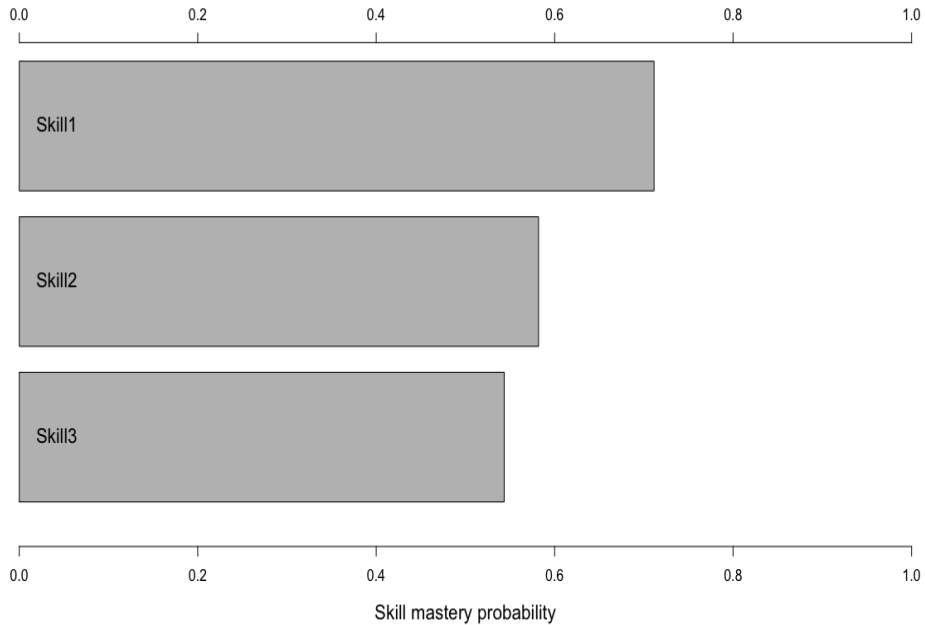
Gráfico 1: Estimación de parámetros para cada ítem de la prueba



Fuente: Elaboración propia con el paquete CDM.

El segundo gráfico muestra la proporción de personas examinadas que poseen un atributo en particular. En el gráfico 2 se ilustra esta situación.

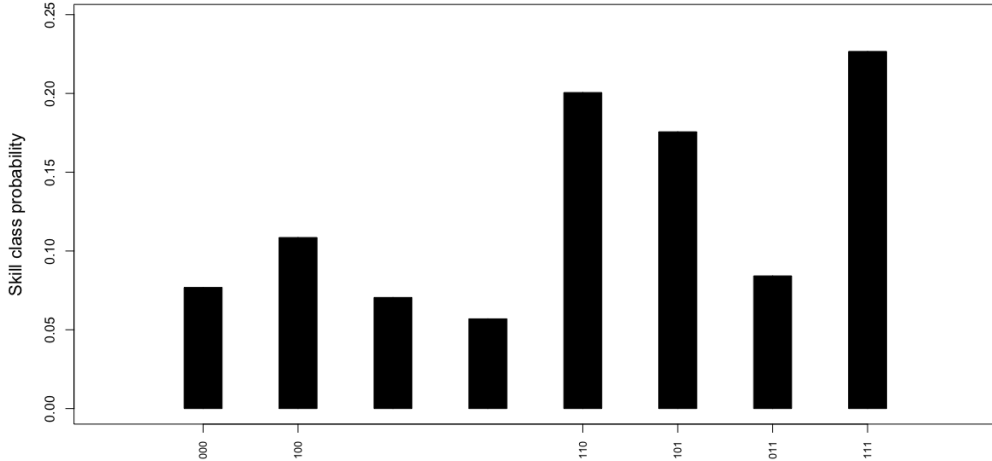
Gráfico 2: Probabilidad de dominio de atributos



Fuente: Elaboración propia con el paquete CDM.

El tercer gráfico representa la proporción de personas que poseen una combinación específica de atributos. Se etiquetan los patrones más frecuentes. Esta situación se ilustra en el gráfico 3.

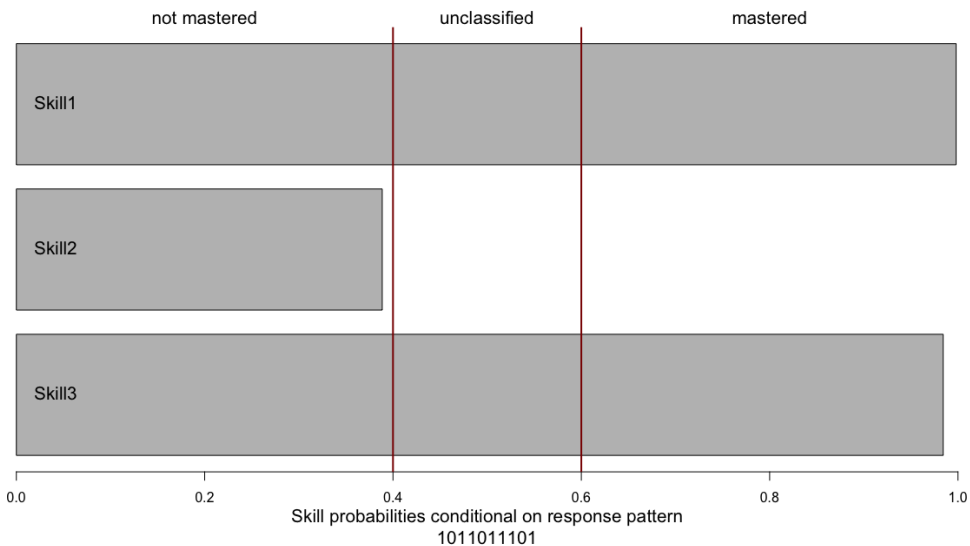
Gráfico 3: Proporción de personas con un perfil específico de atributos



Fuente: Elaboración propia con el paquete CDM

En el cuarto gráfico se muestran los atributos y cada línea vertical representa la probabilidad de dominio del atributo correspondiente. En el gráfico 4 se muestra el perfil individual de atributos para la persona 1, según la clasificación esperada a posteriori, según el modelo DINA.

Gráfico 4: Perfil individual de dominio de atributos para la persona 1



Fuente: Elaboración propia con el paquete CDM

El paquete CDM también incluye un algoritmo para la validación empírica de la matriz Q, el cual se basa en maximizar el índice de discriminación de los ítems. La instrucción para hacer la validación es `din.validate.qmatrix(misdatos)` y, en los resultados del análisis, se puede indicar que no se debe cambiar ningún elemento de la matriz, o bien, se indica cuáles elementos se sugieren modificar. Ante todo, se debe considerar un criterio técnico de especialistas en el área de contenidos que se examina en la prueba, puesto que esta validación únicamente se basa en volver a estimar los parámetros de los ítems y los parámetros de discriminación; en otras palabras, la modificación de una matriz Q debe someterse a juicio de especialistas, quienes dirán si las sugerencias producto del algoritmo son viables.

¿Cuáles *ítems* se ajustan al modelo propuesto? La respuesta a esta interrogante se calcula con la instrucción `misdatos$itemfit.rmsea`. Kunina-Habenicht, Rupp y Wilhelm (2009) recomiendan evaluar al ajuste de los ítems al modelo propuesto de la siguiente manera: los ítems con RMSEA [*root mean square error of approximation*] inferior a .05 indican un buen ajuste; aquellos con valores por debajo de .10 muestran un ajuste moderado y los que superan el .10 tienen un ajuste pobre. Los autores indican que es una regla de dedo para la clasificación. En la tabla 7 se muestra el valor del estadístico para cada ítem y su respectiva clasificación.

Tabla 7: Ajuste de *ítems* al modelo

Ítem	RMSEA	Ajuste
1	.01	bueno
2	.01	bueno
3	.01	bueno
4	.05	bueno
5	.02	bueno
6	.16	pobre
7	.09	moderado
8	.03	bueno
9	.07	moderado
10	.10	pobre

Fuente: Elaboración propia con el paquete CDM

De acuerdo con los resultados de la tabla 7, conviene revisar los ítems 6 y 10, por su pobre ajuste al modelo propuesto. Asimismo, se había indicado anteriormente que el ítem 6 es el que muestra la peor discriminación y resultó fácil para la población examinada. En el gráfico 1 se muestra que tanto el ítem 6 como el 10 son los que tienen mayor probabilidad de acierto por adivinación.

### 3. Discusión

En las últimas dos décadas, el campo de la medición se ha expandido continuamente mediante la investigación científica hacia modelos de diagnóstico cognitivo, los cuales brindan un marco de referencia para la interpretación de resultados de tal manera que el desempeño en una prueba se pueda vincular con inferencias específicas acerca del conocimiento y las habilidades de las personas.

El diagnóstico cognitivo basado en pruebas involucra la especificación y estimación de parámetros y el trabajo con variables latentes de una forma en que los modelos tradicionales de medición no lo requerían. Sin embargo, una de las mayores limitaciones al investigar con modelos de diagnóstico cognitivo ha sido la carencia de software para llevar a cabo los análisis. Este faltante ocasionó que se tuvieran que utilizar varias aplicaciones a la vez, muchas de las cuales tienen un costo económico excesivo o demandan grandes conocimientos técnicos por parte de quienes desean investigar en esta área.

En la actualidad cada vez es mayor la cantidad de códigos que se han desarrollado para funcionar con el entorno R, el cual es muy eficiente en la estimación de parámetros, es gratuito y permite analizar grandes conjuntos de datos y en diversidad de modelos con unas cuantas líneas de sintaxis.

En este artículo se ilustraron algunos usos del paquete CDM, sin embargo, todo su potencial no fue explorado, pues ofrece grandes posibilidades para el desarrollo de algoritmos que permiten interpretaciones más profundas de los datos e incluso, la utilización de modelos más complejos para el diagnóstico cognitivo.

Para futuras investigaciones, se propone una valoración comparativa de este con otros paquetes de R elaborados para tal fin, así como ahondar en el ajuste de los modelos a los datos, la validación empírica de la matriz Q y su uso en estructuras complejas de datos, tales como diseños de matrices múltiples. Finalmente, se recomienda continuar con los esfuerzos de presentar más artículos e investigaciones en idioma castellano, sobre todo para extender el conocimiento de estos modelos y su aplicación en contextos latinoamericanos.

## Referencias

- Artavia-Medrano, A. (2014). *Evaluación cognitiva diagnóstica en Matemática: modelo elaborado con el método rule space para estudiantes costarricenses de undécimo año*. Disertación doctoral no publicada, Universidad de Costa Rica, San José, Costa Rica.
- Chen, Y., Gorin, J., Thompson, M. & Tatsuoka, K. (2008). An alternative examination of Chinese Taipei mathematics achievement: Application of the rule-space method to TIMSS 1999 data. En IEA-ETS Research Institute (Eds.), *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments* (Vol. 1, pp. 23-49). Recuperado el 18 de marzo del 2012, de [http://www.ierinstitute.org/IERI\\_Monograph\\_Volume\\_01.pdf](http://www.ierinstitute.org/IERI_Monograph_Volume_01.pdf)
- Corter, J. & Tatsuoka, K. (2002). *Cognitive and Measurement Foundations of Diagnostic Assessments in Mathematics*. Technical Report of the College Board.
- de la Torre, J. & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333-353.
- Gierl, M. (2007). Making Diagnostic Inferences About Cognitive Attributes Using the Rule-Space Model and Attribute Hierarchy Method. *Journal of Educational Measurement*, 44(4), 325-340.
- Gierl, M., Leighton, J., Wang, C., Zhou, J., Gokiert, R. & Tan, A. (2009). *Validating Cognitive Models of Task Performance in Algebra on the SAT*. Nueva York: The College Board.
- Gierl, M., Roberts, M., Brito, C. & Gotzmann, A. (2009, abril). *Using Judgments from Content Specialists to Develop Cognitive Models for Diagnostic Assessments*. Artículo presentado en Annual Meeting of the National Council on Measurement in Education (NCME), San Diego, CA.
- Im, S. & Yin, Y. (2009). Diagnosing skills of statistical hypothesis testing using the Rule Space Method. *Studies in Educational Evaluation*, 35, 193-199.
- Kunina-Habenicht, O., Rupp, AA. & Wilhelm, O (2009). A Practical Illustration of Multidimensional Diagnostic Skills Profiling: Comparing Results From Confirmatory Factor Analysis and Diagnostic Classification Models. *Studies in Educational Evaluation*, 35, 64-70.

- Leighton, J. & Gierl, M. (2011). *The Learning Sciences in Educational Assessments: The Role of Cognitive Models*. Nueva York: Cambridge University Press.
- Leighton, J., Gierl, M. & Hunka, S. (2002, abril). *The attribute hierarchy model for cognitive assessment*. Louisiana: Artículo presentado en Annual Meeting of the National Council on Measurement in Education (NCME). Recuperado el 9 de setiembre del 2007, de <http://www.education.ualberta.ca/educ/psych/crame>
- Ma, W. & de la Torre, J. (2018). *GDINA: The generalized DINA model framework*. R package version 2.1 <https://CRAN.R-project.org/package=GDINA>
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45, 35-44.
- National Research Council, NRC. (2001). *Knowing what students know: the science and design of educational assessment*. Washington: National Academy Press.
- Norris, S., Leighton, J. & Phillips, L. (2004). What is at stake in knowing the content and capabilities of children's minds?: A case for basing high stakes tests on cognitive models. *Theory and Research in Education*, 2(3), 283-308.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org>
- Ravand, H. & Robitzsch, A. (2015). Cognitive Diagnostic Modeling Using R. *Practical Assessment, Research & Evaluation*, 20(11), 1-12.
- Roberts, M., Brito, C., Gotzmann, A., & Gierl, M. (2009, abril). *Development of Cognitive Models in Mathematics to Promote Diagnostic Inferences about Student Performance*. Artículo presentado en Annual meeting of the American Educational Research Association, San Diego, CA.
- Robitzsch, A., Kiefer, T., George, A. C. & Uenlue, A. (2018). CDM: Cognitive diagnosis modeling. R package version 7.0-12. <https://CRAN.R-project.org/package=CDM>
- Roussos, L. A., DiBello, L. V., Henson, R. A., Jang, E. & Templin, J. (2010). Skills Diagnosis for Education and Psychology With IRT-Based Parametric Latent Class Models. En S. E. Embretson (Ed.) *Measuring Psychological Constructs: Advances in Model-Based Approaches*. Washington: American Psychological Association.

- Rupp, A. A., Templin, J. & Henson, R. A. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. Nueva York: The Guilford Press.
- Rupp, A.A. & van Rijn, P.W. (2018). GDINA and CDM Packages in R. *Measurement: Interdisciplinary Research and Perspectives*, 16(1), 71-77.
- Svetina, D., Gorin, J., Tatsuoka, K. (2011). Defining and Comparing the Reading Comprehension Construct: A Cognitive-Psychometric Modeling Approach. *International Journal of Testing*, 11(1), 1-23.
- Tatsuoka, K. (2009). *Cognitive Assessment: An Introduction to the Rule Space Method*. Nueva York: Routledge Taylor & Francis Group.
- Tatsuoka, K., Guerrero, A., Corter, J., Tatsuoka, C., Yamada, T., Xin, T., et al. (2006). International comparisons of Mathematical Thinking Skills in the TIMSS-R. *Japanese Journal of Research on Testing*, 2(1), 3-40

