**El uso de los Modelos de Diagnóstico Cognitivo para la mejora educativa: Revisión del Estado del arte.**

Adriana Felisa Chávez De la Peña ;)

**Resumen**

Los modelos de diagnóstico cognitivo (CDM) son modelos psicométricos desarrollados para identificar con detalle, a partir de la aplicación de una prueba y a lo largo del conjunto de conocimientos y habilidades cognitivas que conforman el dominio evaluado, las fortalezas y debilidades de cada sustentante, para permitir así el trazo de perfiles detallados que orienten el desarrollo de acciones de mejora focalizadas. La particularidad con que los CDM permiten desglosar el desempeño de los participantes para identificar sus fortalezas y debilidades con detalle, los han convertido en herramientas fundamentales para el desarrollo de diagnósticos específicos en Psicología Clínica, para la selección de personal y el diseño de programas de capacitación y, por supuesto, como instrumentos guía para la mejora educativa. El presente trabajo presenta una introducción completa al mundo de los CDM, su funcionamiento, características y variaciones; así como una revisión de sus principales contribuciones a la Psicometría actual y aplicaciones al mundo real.

**Introducción:**

Existe una línea de investigación en psicometría que se caracteriza por medir múltiples dimensiones o componentes del constructo latente, abandonando la perspectiva unidimensional adoptada por la Teoría Clásica de los Test y la Teoría de Respuesta al Ítem, paradigmas predominantes en Psicometría que interpretan el desempeño observado como reflejo de la posición que ocupan los participantes a lo largo de un solo continuo que representa el dominio de una única variable o habilidad latente evaluada, (de la Torre, 2009). Los nuevos modelos, llamados modelos de diagnóstico cognitivo (CDM por sus siglas en inglés), interpretan el desempeño de los participantes a través de una serie de sub-componentes binarios del dominio general evaluado, para hacer inferencias acerca de cuáles de las habilidades, operaciones cognitivas o conocimientos que componen a este se dominan, y cuáles hace falta reforzar, a partir de los aciertos y errores específicos observados.

De manera general, los CDM están orientados al desarrollo de diagnósticos particularizados que segregan la información recuperada del desempeño de cada sustentante en inferencias detalladas sobre el dominio de distintas operaciones cognitivas. De esta forma, los CDM permiten a todo agente interesado en la aplicación de la prueba, conocer las áreas de fortaleza y debilidad identificadas para orientar el desarrollo de estrategias de mejora. En el ámbito educativo, la aplicación de estos modelos permite a padres de familia, docentes, directores y otros agentes educativos, tomar mejores decisiones con respecto al diseño, uso, manejo e implementación de recursos y materiales didácticos de apoyo hacia los estudiantes, (véase Ketterlin-Geller & Yovanoff, 2009 ó Pérez-Morán, Vázquez-Lira& Rojas, 2019, para un ejemplo)

Para que una evaluación pueda ser considerada para la elaboración de un diagnóstico cognitivo, el diseño de la misma debe considerar la integración de diversas teorías de aprendizaje, cognición y pedagogía y las teorías de medición que permitan el desarrollo de evaluaciones que no se limiten a medir y evaluar, sino que sirvan como insumo para impulsar la mejora del aprendizaje de los estudiantes (Chudowsky y Pellegrino, 2003; Shepard, 2000).

Los modelos de diagnóstico cognitivo requieren de una matriz (identificada en la literatura como matriz Q), que identifique para cada ítem, cuáles son las operaciones cognitivas que se consideran “necesarias” para obtener un acierto. Su construcción requiere del trabajo conjunto de expertos en el dominio evaluado, sustentantes que participen en un piloteo activo y reflexivo que permita identificar los procedimientos seguidos y de psicómetras que estén constantemente cotejando la información recopilada a la luz de las respuestas observadas.

**Acerca de las aplicaciones de los CDM**

La particularidad y el detalle con que los CDM permiten obtener información sobre los distintos componentes que se sabe forman parte del dominio de interés, los han vuelto herramientas altamente valoradas y utilizadas a lo largo de distintas áreas. Por ejemplo, en Psicología Clínica (donde muchas veces no es posible medir los constructos asociados a la salud mental de otra forma que no sea con la aplicación de un instrumento psicológico), se ha aprovechado la mirada segregada de los CDM para estudiar y tipificar las características de desórdenes de la personalidad (Rossi, Elkit & Simonsen, 2010; Templin & Henson, 2006); para estudiar algunas enfermedades mentales como la Ezquisofrenia (Jaeger, Tatsuoka & Berns, 2003).

**Los CDM y la mejora educativa**

Una de las áreas que más beneficio han obtenido de la aplicación de CDMs al análisis e interpretación de las respuestas observadas en una prueba, es el ámbito Educativo en general. La identificación particular de áreas de fortaleza y debilidad constituye un elemento invaluable para la mejora educativa.

A este respecto, resaltan las aplicaciones de los CDMs al ámbito de las matemáticas (Brown y Burton, 1978; Chen y Macdonald, 2011; Gierl, Leighton, Changjiang, Jiawen, Rebecca & Tan, 2009; Ma, Çetin y Green, 2009; Pérez-Morán, 2014; Pérez-Morán, Contreras-Roldan, Hernández, Olivares, Chan, y Díaz, 2014; Pérez-Morán, Larrazolo, Backhoff, y Guaner, 2015; Revuelta y Ponsoda, 1998; Romero, Ponsoda y Ximénez, 2008; Birenbaum & Tatsuoka, 1993 ; Pérez-Morán, Vázquez-Lira, Rojas, 2019; Ketterlin-Geller & Yovanoff, 2009), la comprensión lectora (Lee & Sawaki, 2009, Jang, 2009; Li, 2011; Li, Hunter & Lei, 2016; Ravand, 2016) y algunas otras evaluaciones generales (Montero, Monfils, Wang, Yen Julian & Moody, 2003)

**Revisión de los distintos CDM en la literatura**

Existe una amplia gama de modelos de diagnóstico cognitivo (CDM), (para una revisión extensa, se recomienda revisar Cohen, (2019), Ferrara, Lai, Reilly, Nichols, Rupp y Leighton, (2017), o a van der Linden, (2016). Cada CDM hace supuestos específicos acerca de cómo se pondera el dominio de las distintas operaciones cognitivas para producir un acierto o error. Por ejemplo, una distinción importante tiene que ver con si el modelo es conjuntivo o disyuntivo (Rupp, Templin y Henson, 2010): los modelos conjuntivos asumen que se requiere el dominio de todos los atributos asociados al ítem para poder obtener un acierto, mientras que los modelos disyuntivos asumen que la falta de dominio de un atributo puede ser compensada por el dominio de otros atributos.

Algunos de los CDMs más desarrollados y utilizados en la literatura son el modelo conjuntivo DINA (entrada determinística, ruidosa "y" puerta; Junker & Sijtsma, 2001; de la Torre, 2009) y su variante disyuntiva, el modelo DINO (entrada determinística, ruidosa "o" puerta; Templin y Henson, 2006), y el A-CDM (CDM aditivo; de la Torre, 2011). Según Rupp, Templin y Henson, (2010), otros CDM bien conocidos son el modelo NIDA (determinista de entrada ruidosa “y”; Junker y Sijtsma, 2001, Maris, 1999), el NIDO (determinista de entrada ruidosa o, Douglas, de la Torre, Chang, Henson & Templin, 2006), y el R-RUM (modelo unificado de reparación reducida; Hartz, 2002). Además, se han propuesto CDM generales que respetan los supuestos de modelos específicos (véase, por ejemplo, Henson, Templin & Willse, 2009; von Davier, 2005), o bien, que funcionan como modelos generales, como es el caso del modelo G-DINA (DINA generalizada; de la Torre, 2011), el modelo de diagnóstico cognitivo log-lineal (LCDM; Henson, Templin y Willse, 2009) y el modelo de diagnóstico general (GDM; von Davier , 2005). Este último grupo de modelos, describe la probabilidad de éxito en términos de la suma de los efectos debidos a la presencia de atributos específicos y sus interacciones.

**Más acerca del Modelo DINA**

El modelo DINA constituye uno de los modelos más sencillos dentro de la familia de los CDMs (Junker y Sijtsma, 2001; de la Torre, 2009), pues considera únicamente dos parámetros libres por ítem, con el objetivo de describir la probabilidad de que los aciertos o errores registrados no estén relacionados con el grado de dominio que los sustentantes tienen en las habilidades requeridas. Estos parámetros, conocidos como parámetro de adivinación y parámetro de desliz, refieren a la probabilidad de obtener un acierto aún sin dominar las habilidades necesarias, (es decir, de “atinarle por por puro azar” a la respuesta correcta), y la probabilidad de errar el ítem aún dominando las habilidades necesarias, (en otras palabras, de cometer un “desliz” al momento de seleccionar una respuesta).

El modelo DINA se expresa a partir de la siguiente ecuación,

De acuerdo con la Ecuación 1, el modelo DINA está compuesto por los siguientes elementos:

1.- Una variable binaria para indicar si la persona obtuvo un acierto (1) o un error (0) en el ítem . Esto implica la existencia de un vector Y compuesto por filas e columnas, donde se concentran los aciertos y errores obtenidos por cada persona en cada ítem.

2.- Un vector que contiene a la variable binaria que señala si la persona domina (1) o no (0) el subdominio evaluado en la prueba. Al conjunto de dominios evaluados suele referírseles, en el marco de los CDM, como habilidades, conocimientos o atributos contenidos en el dominio general evaluado por la prueba.

3.- Un vector que por cada ítem , contiene a la variable binaria que señala la respuesta “idealmente” registrada (en términos de acierto o error) por el sustentante en el ítem , dado lo que la matriz Q, el modelo cognitivo detrás de la prueba, ha establecido acerca de qué habilidades se requieren para obtener un acierto en cada ítem ( y lo que el vector nos dice sobre las habilidades dominadas por el participante . De esta forma, tenemos que el vector queda plasmado como el elemento determinista del modelo, definido matemáticamente como:

, donde

4.- Un parámetro de desliz , que nos dice que aunque la respuesta idealmente esperada por el participante al ítem , sea 1 (porque domina todas las habilidades requeridas por el ítem, ), el participante cometa un “desliz” y se observe aun así que . Es decir,

)

5.- Un parámetro de adivinación , que representa la probabilidad de que un examinado que no posee todas las habilidades requeridas por el ítem , (), “le atine” a la respuesta correcta y respondan correctamente el ítem (). Es decir,

El modelo DINA nos dice que para cada ítem , solo los examinados que dominen todos los atributos requeridos tendrán una probabilidad de éxito igual a , mientras que todos los demás examinados tendrán una probabilidad de éxito igual a . Es decir, que como habíamos mencionado previamente, el modelo DINA es un modelo conjuntivo que asume que es necesario el dominio de todos y cada uno de los atributos relacionados con cada ítem.

**Técnicas de retrofitting**

Como ya se ha mencionado, idealmente para que una prueba pueda ser utilizada para la elaboración de un diagnóstico cognitivo, esta debería haber sido diseñada explícitamente con esa intención, incorporando para ello técnicas y conocimientos provenientes de la Pedagogía, la Psicometría y de toda disciplina que guarde relación con el dominio que interesa evaluar. Sin embargo, existen ocasiones en que la aplicación de un CDM a los resultados obtenidos tras la aplicación de una prueba a gran escala puede considerarse altamente valiosa para obtener un diagnóstico de alto impacto, que pueda arrojar información valiosa acerca de toda una población. Para ello, se requiere de la aplicación de una serie de técnicas y métodos (conocidas en la literatura como “técnicas de retrofitting”) que permitan identificar el conjunto de operaciones cognitivas evaluadas en la prueba.

Como un ejemplo, tómese el estudio realizado por Pérez-Morán, Vázquez-Lira y Rojas, (2019), quienes aplicaron técnicas de retrofitting sobre la Prueba de Matemáticas para primaria (06) del PLANEA ELCE 2015 (INEE, 2015), aplicada a gran escala a los estudiantes de sexto año de primaria de México, para elaborar un diagnóstico nacional del dominio que tienen los estudiantes del país a lo largo de las habilidades básicas en matemáticas evaluadas en la prueba. Como parte de este estudio, los autores realizaron estudios cognitivos para identificar las habilidades cognitivas requeridas por los distintos ítems que conforman la prueba, mediante la aplicación de técnicas de retrofitting. Otros ejemplos pueden encontrarse en Jang, (2009) ó Li, (2011)

Toda técnica de retrofitting requiere comenzar con la realización de estudios cognitivos que permitan identificar los modelos, estrategias y procesos de respuesta adheridos a cada ítem contenido en la prueba. Esto obedece a la necesidad señalada por autores como Yang y Embretson, (2007) de que toda prueba de diagnóstico cognitivo diseñada con propósitos de mejora, deben ser diseñadas y validadas a partir de modelos cognitivos que permitan identificar de manera confiable los procesos de respuesta asociados a cada ítem. Para ello, es deseable revisar la genealogía, la congruencia y alineación de los ítems.

Es recomendable trabajar con un grupo de expertos que dominen los principales contenidos sustantivos asociados al dominio general evaluado, la enseñanza de los mismos (en el caso de pruebas con impacto académico) y que tengan conocimientos acerca de la construcción de instrumentos de medida (Rupp, Templin y Henson, 2010).

Algunos autores sugieren que la prueba a evaluar sea revisada en términos de los elementos de análisis del modelo para la Evaluación del Diseño Universal (EDU) propuestos por Thompson, Johnstone y Thurlow (2002), que ha demostrado ser de gran utilidad para el desarrollo de evaluaciones más accesibles para los examinados (Johnstone, 2003) y para minimizar la varianza irrelevante del constructo originada por problemas en el diseño, formato y sesgos culturales presentes en los ítems (Haladyna, Downing, y Rodríguez, 2002).

Es imperante realizar un piloteo de los ítems contenidos en la prueba para la obtención de reportes verbales exhaustivos de los procesos de respuesta empleados, (Ericsson & Simon, 1993). Idealmente, el piloteo no sólo se realizará con un grupo de sustentantes que emulen las características de la población de interés, sino también con un grupo de expertos en el dominio evaluado. La obtención de reportes verbales debe realizarse de manera sistematizada, a partir del uso de protocolos de pensamiento en voz alta con técnicas concurrentes y retrospectivas (Ericsson y Simon, 1984, 1993; Leighton, 2009; Leighton y Gierl, 2007). Como análisis complementarios, se recomienda la aplicación de análisis del sendero de la vista (*eye-tracking*; Snow y Lohman, 1989; Sternberg, 1977), así como la revisión de las latencias de respuesta (Fredericksen, 1980; Posner, 1978; Posner y Rogers, 1978). Estas medidas complementarias, se consideran de gran ayuda para obtener información en los casos en que se presentan dificultades para evocar el reporte verbal de los procedimientos seguidos por estudiantes de corta edad, o bien, de procedimientos que suceden en cuestión de tan sólo algunos segundos y que, por tanto, no es posible su introspección (Sternberg, 1977), y su aplicación permite una mejor verificación de la relación entre el modelo cognitivo elaborado por los expertos y los procesos cognitivos utilizados y reportados por los examinados para responder los ítems (Messick, 1989).

La información recuperada de los estudios ya mencionados debe revisarse a consciencia para lograr identificar e integrar los modelos cognitivos asociados a los diversos procesos de respuesta y desglosarlos en operaciones cognitivas concretas que permitan un mayor nivel explicativo del comportamiento de cada ítem.

Finalmente, para valorar la pertinencia de aplicar un CDM a los resultados obtenidos tras la aplicación de la prueba de interés, es necesario corroborar que no hay unidimensionalidad en términos del constructo evaluado. Para ello, se recomienda realizar análisis factoriales exploratorios, para descartar la existencia de un solo factor dominante, (que estaría más en acuerdo con los supuestos planteados por la TCT y la TRI).

**Referencias**

* Birenbaum, M., & Tatsuoka, K. K. (1993). Applying an IRT-based cognitive diagnostic model to diagnose students' knowledge states in multiplication and division with exponents. *Applied measurement in education*, *6*(4), 255-268.
* Brown, J. & Burton, R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 2, 155-192.
* Chen, Y. & Macdonald, G. (2011). Validating Cognitive Sources of Mathematics Item Difficulty: Application of the LLTM to Fraction Conceptual Items. *Psychological Assessment*, *7*, 74–93.

Chudowsky, N., & Pellegrino, J. W. (2003). Large-scale assessments that support learning: What will it take?. *Theory into practice*, *42*(1), 75-83.

Cohen, Y. (2019). The Handbook of Cognition and Assessment; Frameworks, Methodologies, and Applications.

* De La Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of educational and behavioral statistics*, *34*(1), 115-130

De La Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179-199.

Douglas, J., de la Torre, J., Chang, H., Henson, R., & Templin, J. (2006, April). Skills diagnosis with latent variable models. In *annual meeting of the National Council on Measurement in Education, San Francisco, CA*.

* Ericsson, K. & Simon, H. (1984). *Protocol analisys: verbal reports as data*. Cambridge: MIT Press.
* Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT.

Ferrara, S., Lai, E., Reilly, A., Nichols, P. D., Rupp, A. A., & Leighton, J. P. (2017). Principled approaches to assessment design, development, and implementation. *The Handbook of Cognition and Assessment*, 41-74.

* Fredericksen, J. (1980). Component skills in Reading: measurements of individual diferences thought chronometric analisys. In R. E. Snow, P-A. Federico & W. E. Montage (Eds.), *Aptitude, learning, and instructions: Cognitive process analyses of aptitude*, Vol. 1, (pp. 105-138). Hillsdale: Lawrence Erlbaum.
* Gierl, M., Leighton, J., Changjiang, W., Jiawen, Z., Rebecca, G. & Tan, A. (2009). *Validating Cognitive Models of Task Performance in Algebra on the SAT. Research Report 2009-3*. College Board, Research Report, 2009(3). New York.
* Haladyna, T. Downing, S. M. & Rodríguez, M. C. (2002). A review of multiple-choice item writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*(3), 309–334.

Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality*(Doctoral dissertation, ProQuest Information & Learning).

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191.

* INEE. (2015). Plan Nacional para la Evaluación de los aprendizajes (PLANEA). Resultados nacionales 2015. Recuperado de http://www.inee.edu.mx/images/stories/2015/planea/inal/fascículosinales/

resultadosPlanea-3011.pdf

Jaeger, J., Tatsuoka, C., & Berns, S. M. (2003). Innovative methods for extracting valid cognitive deficit profiles from NP test data in schizophrenia. Schizophrenia Research, 1(60), 140.

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), 031-73.

* Johnstone, C. (2003). Improving validity of large-scale tests: Universal design and student performance (Technical Report 37). Minneapolis: National Center on Educational Outcomes.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258-272.

Ketterlin-Geller, L. R., & Yovanoff, P. (2009). Diagnostic assessments in mathematics to support instructional decision making. *Practical Assessment, Research & Evaluation*, *14*(16), 1-11.

Lee, Y. W., & Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly*, *6*(3), 239-263.

* Leighton, J. & Gierl, M. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees’ thinking processes. *Educational Measurement: Issues and Practice, 26*(2), 3-16.
* Leighton, J. (2009). Two Types of Think Aloud Interviews for Educational Measurement: Protocol and Verbal Analysis Paper presented for symposium How to Build a Cognitive Model for Educational Assessments at the 2009 annual meeting of the National Council on Measurement in Education (NCME), April, 14-16.
* Li, H. (2011). A cognitive diagnostic analysis of the MELAB reading test. *Spaan Fellow*, *9*, 17-46.
* Li, H., Hunter, C. V., & Lei, P. W. (2016). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, *33*(3), 391-409.
* Ma, L. Çetin, E. y Green, K. (2009, April). *Cognitive assessment in Mathematics with the Least Squares Distance Method.* Artículo presentado en el Congreso anual de la AERA 2009. San Diego.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*(2), 187-212.

* Messick, S. (1989). Validity. En R. L. Linn (Ed.), *Educational measurement* (3a. ed.), pp. 13-103. New York: Macmillan Publishing Co.
* Montero, D. H., Monfils, L., Wang, J., Yen, W. M., Julian, M. W., & Moody, M. (2003, April). Investigation of the application of cognitive diagnostic testing to an end-of-course high school examination. In *annual meeting of the National Council on Measurement in Education, Chicago, IL*.
* Pérez-Morán, J. C. (2014). *Análisis del aspecto sustantivo de la validez de constructo de una prueba de habilidades cuantitativas* (tesis doctoral). Universidad Autónoma de Baja California, Baja California, México.
* Pérez-Morán, J. C., Contreras, S., Hernández, E. M., Olivares, C., Chan, P., y Díaz, K. M. (2014). Análisis de las evidencias de validez basadas en el proceso de respuesta de las pruebas de ENLACE MS de Habilidad lectora y Matemáticas. Reporte técnico. México: INEE
* Pérez-Morán, J. C.; Larrazolo, N.; Backhoff, E.; y Guaner, R. (2015). Análisis de la estructura cognitiva del área de habilidades cuantitativas del EXHCOBA mediante el modelo LLTM de Fisher. Revista Internacional de Educación y Aprendizaje, 3(1), 25-38. <http://coleccionderevistasdeeducacionyaprendizaje.cgpublisher.com/product/pub.329/prod.5> ISSN 2255-453X
* Pérez-Morán, J. C.; Vázquez-Lira, R.; & Rojas, G. (2019). Diagnóstico Nacional de las habilidades básicas en Matemáticas de Sexto de Primaria: Resultados de 2015. México: RIMEDIE.
* Posner, M. I. (1978). *Chronometric exploration of mind.* New York: Jhon Wiley.

Posner, M. I., & Rogers, M. G. K. (1978). Chronometric analysis of abstraction and recognition. In W. K. Estes (Ed.) (1978). *Handbook of learning and cognitive processes* (vol. 6). Hillsdale, N. J.: Lawrecence Erlbaum Associates.

Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*, *34*(8), 782-799.

Revuelta, J. y Ponsoda, V. (1998). Un test adaptativo informatizado de análisis lógico basado en la generación automática de ítems. *Psicothema, 10*, 753-760.

Romero, S., Ponsoda, V., y Ximenez, C. (2008). Análisis de un test de aritmética mediante el modelo logístico lineal de rasgo latente 1. *Revista Latinoamericana de Psicología, 40*, 85–95.

Rossi, G., Elklit, A., & Simonsen, E. (2010). Empirical Evidence for a Four Factor Framework of Personality Disorder Organization: Multigroup Confirmatory Factor Analysis of the Millon Clinical Multiaxial Inventory—III Personality Disorder Scales Across Belgian and Danish Data Samples. *Journal of Personality Disorders*, *24*(1), 128-150.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). Diagnostic assessment: Theory, methods, and applications. *New York: Guilford*.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational researcher*, *29*(7), 4-14.

* Snow, R. & Lohman, D. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3a. ed.), pp. 263-331. New York: Macmillan Publishing Co.
* Sternberg, R. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Oxford: Lawrence Erlbaum.
* Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods*, *11*(3), 287.

Templin, J., & Henson, R. A. (2006). A Bayesian method for incorporating uncertainty into Q-matrix estimation in skills assessment. In *Symposium conducted at the meeting of the American Educational Research Association, San Diego, CA*.

* Thompson, S., Johnstone, C. & Thurlow, M. (2002). *Universal design applied to large scale assessments (Synthesis Report 44).* Minneapolis, MN: National Center on Educational Outcomes.
* Van der Linden, W. J. (Ed.). (2017). *Handbook of Item Response Theory, Volume Three: Applications*. CRC Press.

von Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report Series*, *2005*(2), i-35.

Willse, J. T., & Shu, Z. (2014). CTT: Classical test theory functions. *R package version*, *2*.

* Yang, X. & Embretson, S. (2007). Construct Validity and Cognitivy Diagnostic Assesment. In Leighton, J. y Griel, M. (Edit.). *Cognitive diagnostic assessment for education: Theory and applications*, pp. 85-118. Cambridge: Cambrige University Press.