



EL MÉTODO DE JERARQUÍA DE ATRIBUTOS: UNA OPCIÓN PARA VINCULAR TEORÍAS COGNITIVAS Y DEL APRENDIZAJE CON PRÁCTICAS PSICOMÉTRICAS Y DE ENSEÑANZA

ATTRIBUTE HIERARCHY METHOD: AN ALTERNATIVE TO LINK COGNITIVE AND
LEARNING THEORIES WITH PSYCHOMETRIC AND TEACHING PRACTICES

Volumen 11, Número 1
pp. 1-29

Este número se publicó el 30 de abril de 2011

Álvaro Artavia Medrano

La revista está indexada en los directorios:

[LATINDEX](#), [REDALYC](#), [IRESIE](#), [CLASE](#), [DIALNET](#), [DOAJ](#), [E-REVIST@S](#),

La revista está incluida en los sitios:

[REDIE](#), [RINACE](#), [OEI](#), [MAESTROTECA](#), [PREAL](#), [HUASCARAN](#), [CLASCO](#)

Los contenidos de este artículo están bajo una licencia [Creative Commons](#)



EL MÉTODO DE JERARQUÍA DE ATRIBUTOS: UNA OPCIÓN PARA VINCULAR TEORÍAS COGNITIVAS Y DEL APRENDIZAJE CON PRÁCTICAS PSICOMÉTRICAS Y DE ENSEÑANZA

ATTRIBUTE HIERARCHY METHOD: AN ALTERNATIVE TO LINK COGNITIVE AND LEARNING THEORIES WITH PSYCHOMETRIC AND TEACHING PRACTICES

Álvaro Artavia Medrano¹

Resumen: Dado que en la mayoría de las pruebas en el ámbito educativo las personas se involucran en algún tipo de resolución cognitiva de problemas, los enfoques psicométricos tradicionales tienen el inconveniente de no aportar información acerca de las estructuras de conocimiento ni de las habilidades subyacentes a la resolución de los ítems de una prueba. Por ello, se necesita de enfoques de evaluación que se orienten a aspectos cognitivos y que permitan diseñar pruebas e interpretar resultados en términos de los procesos y las habilidades propias de la resolución de problemas. En este artículo se presenta una de las opciones más recientes de la evaluación cognitiva diagnóstica: el Método de Jerarquía de Atributos (MJA), que puede ser visto como un modelo cognitivo de desempeño que resume un conjunto de conocimientos y habilidades interconectadas que las personas utilizan para responder correctamente los ítems de una prueba en un dominio específico. Este enfoque posibilita contar con métodos de evaluación orientados al mejoramiento del aprendizaje y proponer diversas técnicas de enseñanza basadas en la psicología del desempeño en la resolución de tareas.

Palabras clave: EVALUACIÓN COGNITIVA DIAGNÓSTICA, MÉTODO DE JERARQUÍA DE ATRIBUTOS, MODELOS COGNITIVOS

Abstract: Since most of educational tests require examinees to engage in some form of cognitive problem solving, traditional psychometric approaches have the disadvantage of not provide information about knowledge structures nor underlying skills to solve test items. Therefore, cognitive assessment approaches are required, which lead to design tests and to interpret results in terms of process and skills related to problem solving. This article presents one of the most recent options of cognitive diagnostic assessment: the Attribute Hierarchy Method (AHM), which can be viewed as a cognitive model of task performance; it summarizes the set of interconnected knowledge and skills that students use to respond correctly to test items within a specific domain. This approach allows for methods of assessment aimed to improve learning and to propose teaching situations based on the psychology of performance in problem solving.

Key words: COGNITIVE DIAGNOSTIC ASSESSMENT, ATTRIBUTE HIERARCHY METHOD, COGNITIVE MODELS

¹ Candidato a Doctor en Educación del Programa Latinoamericano de Doctorado en Educación de la Universidad de Costa Rica (UCR). Actualmente, es profesor de la Escuela de Formación Docente de la Universidad de Costa Rica y funcionario del Ministerio de Educación Pública de Costa Rica.

Direcciones electrónicas: alvaro.artavia@ucr.ac.cr,
alvartavia@gmail.com

Artículo recibido: 19 de noviembre, 2010

Aprobado: 28 de marzo, 2011

Introducción

Dado que la evaluación tradicional no toma en cuenta la compleja gama de estructuras de conocimiento que las personas utilizan para procesar la información, los marcos de interpretación de los resultados no reflejan un acuerdo profesional ni social sobre qué debe haber aprendido el estudiantado. Tales resultados parecieran considerar que

(...) los estudiantes logran los mismos productos o resultados académicos recorriendo necesariamente unas mismas secuencias de procesos, es decir, enfocando del mismo modo la variedad de temas que estudian, razonando del mismo modo a la hora de tratarlos, demostrando niveles semejantes de habilidad en todos ellos. (Bernad, 2000, p.35)

Aunado a lo anterior, se presentan otros factores tales como: 1) priorización de contenidos que pueden ser medidos con facilidad –no necesariamente los más importantes según los supuestos acuerdos entre los actores del sistema–, 2) falta de comunicación entre las unidades de evaluación y las dedicadas al diseño y el desarrollo curricular nacional –entre otras razones, por el diferente origen profesional de sus integrantes–, y 3) falta de coherencia entre las actividades de evaluación y las metas de aprendizaje propuestas oficialmente (Ferrer, 2006).

No obstante, los avances en el estudio de la cognición, el aprendizaje y el desarrollo humano brindan nuevas perspectivas para el diseño, la administración y el uso de evaluaciones del rendimiento académico, con el propósito de alcanzar una de las principales metas de la educación contemporánea: ayudar al estudiantado a lograr el éxito escolar (Pellegrino, Baxter y Glaser, 1999). Para ello, las prácticas de evaluación y de enseñanza necesitan de una comprensión más detallada de la adquisición de competencias y dominio de áreas específicas de conocimiento, para así tener una mayor y positiva influencia tanto en lo que se aprende en los salones de clase como en lo que se mide en pruebas de aula y de alcance nacional.

De esta manera, en este artículo se asume un concepto distinto de evaluación al utilizado tradicionalmente, pues se le concibe como una parte integral de la enseñanza que contribuye significativamente al aprendizaje de los conocimientos identificados como relevantes y pertinentes para un sistema educativo en particular. Consecuentemente, los logros de aprendizaje se consideran como desarrollos graduales de habilidades cognitivas y, por ello, se necesita que la evaluación ofrezca información útil acerca del cambio en la organización y la estructura del conocimiento, sobre todo explicitada en descripciones de lo que el estudiantado utiliza para tener éxito en el dominio de conocimientos que está aprendiendo.

Este creciente interés, por obtener evidencias sobre los procesos mentales de las personas, ha generado una presión en la evaluación educativa para que brinde información que permita determinar en qué medida las puntuaciones obtenidas en una prueba reflejan ciertas formas de pensamiento y procesos cognitivos de niveles superiores asociados a un aprendizaje significativo.

Autores como Tatsuoka (1983), Messick (1984), Bejar (1984), Frederiksen (1990), Nichols (1994), Pellegrino, Baxter y Glaser (1999), Junker y Sijtsma (2001), Leighton, Gierl y Hunka (2002), Mislevy, Steinberg y Almond (2003) y Gorin (2006), entre otros, han planteado la necesidad de combinar la psicología cognitiva con la psicometría para identificar la forma en que el estudiantado organiza su conocimiento y así diagnosticar las concepciones erróneas que evidencien y, con ello, proponer acciones para el mejoramiento del aprendizaje.

Por lo tanto, el propósito del presente artículo es analizar una de las propuestas en la que se integran teorías cognitivas y del aprendizaje con prácticas psicométricas y de enseñanza, en particular el método de jerarquía de atributos. Esta forma de evaluación cognitiva diagnóstica se propone dada la necesidad de comprender los procesos cognitivos involucrados en la resolución de situaciones en el ámbito educativo, lo que permitirá diseñar experiencias de aprendizaje que promuevan el desarrollo cognitivo en forma conjunta con la adquisición de conocimientos por parte del estudiantado.

El artículo se ha organizado en cinco secciones. En la primera, se presentan conceptos fundamentales de la evaluación cognitiva diagnóstica como una forma de integración entre la psicometría y la psicología cognitiva, así como una discusión de la relevancia de los modelos cognitivos en este tipo de evaluaciones en el campo educativo. En la segunda sección, se trabaja con el concepto de atributos y la posibilidad de estructurarlos jerárquicamente. En la tercera sección, se hacen planteamientos generales acerca del método propuesto por Tatsuoka y que constituye los cimientos para el modelo psicométrico-cognitivo que se estudia con mayor detalle en la cuarta sección. El artículo concluye con una sección de consideraciones generales.

Evaluación cognitiva diagnóstica

Nichols (1994) utilizó el término "evaluación cognitiva diagnóstica" con el propósito de capturar la integración de la psicología cognitiva en el diseño e interpretación de los resultados de evaluaciones, así como en la elaboración de diagnósticos sobre las características del estudiantado en cuanto a sus procesos cognitivos. Por su parte, Rupp (2007) considera que

dicho término tiene algunas restricciones semánticas, pues no permite considerar tendencias desarrolladas recientemente para integrar la psicología cognitiva con la medición, pero que no tienen un propósito diagnóstico (por ejemplo, la generación automatizada de ítems). En consecuencia, Rupp propone los términos "psicometría cognitiva", "modelos psicométrico-cognitivos" e "investigación psicométrico-cognitiva". No obstante, dado el alcance del tratamiento de la información en este artículo, se utilizará el término propuesto por Nichols.

La evaluación cognitiva diagnóstica se ha destinado a la medición de estructuras específicas de conocimiento y habilidades de procesamiento de estudiantes para brindar información sobre sus fortalezas y debilidades cognitivas, así como para mejorar sus oportunidades de aprendizaje.

Para Leighton y Gierl (2007b) las ideas, teorías y métodos de la psicología cognitiva pueden contribuir al avance de la medición educativa, dado que: 1) brindan información sobre análisis de pruebas ya existentes para así dilucidar sus constructos subyacentes, 2) clarifican las metas de la elaboración de pruebas en términos del conocimiento y las habilidades requeridas, como verdaderos indicadores de dominio y entendimiento, y 3) mejoran las teorías de aptitudes, el logro y los aprendizajes en diferentes ámbitos. De esta manera, los aspectos cognitivos subyacentes a las pruebas en el campo educativo son de gran utilidad en la comprensión de los constructos medidos y pueden ser descritos en términos de las estrategias, los procesos y las representaciones del conocimiento que evidencian las personas (Gierl, Leighton y Hunka, 2000).

Al discutir sobre la necesidad y la relevancia de la evaluación cognitiva diagnóstica, el primer plano de la discusión lo ocupa la forma en la que han de adaptarse los modelos propios de la teoría psicológica a la medición educativa, siendo consecuentes con los principios estructurales y fundamentales de tal teoría.

Los modelos de medición psicométrico-educativos, en general, se proponen aproximar la ubicación de una persona respecto de una variable de interés; los basados en la teoría de respuesta al ítem han contribuido, en gran medida, a la medición psicológica y educativa al superar algunos obstáculos técnicos, tales como el hecho de que la estimación de la habilidad de una persona dependa de una muestra particular de los ítems de una prueba. Sin embargo, esta forma de medir ha evidenciado ciertas limitaciones sobre todo en contextos educativos cambiantes en los que cada vez es mayor la demanda de información sobre el procesamiento cognitivo del estudiantado.

Las limitaciones de los modelos de medición psicométrico-educativos se han identificado específicamente en cuanto a su deficiencia al incorporar teoría psicológica sustantiva en la explicación de las respuestas dadas a los ítems y en relación con supuestos basados en la realidad acerca de las dependencias psicológicas y las variables que influyen en el desempeño en un ítem de una prueba particular. De igual forma, otra carencia de los modelos consiste en una clara falta de determinación de los procesos psicológicos que reflejan colectivamente los constructos medidos por una prueba, pues gran parte de lo que informan las pruebas del ámbito educativo son reflejo de las expectativas que quienes investigan tienen sobre cómo las personas razonan y resuelven problemas en una prueba y no se basan en evidencia empírica de cómo tales personas piensan en esas circunstancias (Huff y Goodman, 2007).

Existen teorías cognitivas para muchos fenómenos, tales como percepción, memoria, atención, razonamiento, resolución de problemas, inteligencia y habilidades espaciales. Sin embargo, hay muy poco desarrollo en cuanto a teorías cognitivas relacionadas, en particular, con la evaluación o con los procesos multidimensionales y complejos que pretenden medir algunas pruebas. Por tal motivo, se han incorporado teorías que provienen de la psicología cognitiva en el campo de la medición educativa, pero esto ha resultado de extrema dificultad debido a que el desarrollo de tales teorías se ha dado en contextos reducidos de aprendizaje y en ausencia de marcos teóricos formales de evaluación (Leighton y Gierl, 2007a). De esta manera, se considera una responsabilidad de quienes investigan en el campo educativo, la adaptación de métodos, técnicas y herramientas necesarias para incorporar teorías cognitivas, sobre todo en lo que se refiere a propósitos de medición en la educación.

Las pruebas del ámbito educativo diseñadas con propósitos de diagnóstico cognitivo difieren de las elaboradas según enfoques tradicionales. Estas últimas se basan únicamente en taxonomías lógicas y especificaciones de contenido en la descripción de sus objetivos, de tal manera que sus esfuerzos se concentran en revelar los mecanismos que las personas utilizan al responder ítems. Por su parte, las pruebas del ámbito educativo, diseñadas con propósitos de diagnóstico cognitivo, pretenden apoyarse en la psicología del aprendizaje para describir los procesos de razonamiento y resolución de problemas.

Para cumplir con su cometido, las pruebas desarrolladas con una finalidad de diagnóstico cognitivo deben asegurar cuáles son las estructuras específicas de conocimiento y los procedimientos que se miden. Nichols (1994) ha propuesto cinco pasos:

- 1) Construcción de una teoría sustantiva, esto es, el desarrollo de un modelo o teoría que caracterice las estructuras hipotéticas de conocimiento y los procesos requeridos para el desempeño en una evaluación, por lo que deben identificarse las variables de los ítems que se refieren a estructuras de conocimiento y procesos cognitivos particulares.
- 2) Selección de un diseño, es decir, con la guía del modelo o teoría seleccionado según el paso anterior, quien elabora la prueba debe escoger el diseño de observación y medición, esto es, que los ítems escogidos o creados sean seleccionados de tal manera que las personas respondan a los ítems de formas predecibles, con los procesos cognitivos y las estructuras de conocimiento identificados en el paso 1.
- 3) Administración de la prueba, lo que involucra detalles tanto del entorno como del contexto en que se desarrollará: formato y presentación de los ítems y establecimiento de la prueba en términos generales. Se recomienda que las decisiones en torno a este paso sean informadas para eventualmente determinar su incidencia en el desempeño de las personas.
- 4) Calificación de las respuestas con el propósito de operacionalizar el diseño de evaluación seleccionado.
- 5) Revisión del diseño para así acumular evidencia gradualmente que fundamente la teoría seleccionada o bien, sugiera cambios como producto de los resultados de la evaluación.

De esta manera, al tener presente las orientaciones dadas para su correcta elaboración, se reconoce que la evaluación cognitiva diagnóstica brinda información sobre las estrategias utilizadas por las personas en la resolución de situaciones, las relaciones entre los conceptos que perciben y los principios propios de un dominio evaluado. La especificidad de la información está dada en términos explicativos de por qué el estudiantado respondió de una manera determinada, es decir, cómo vincular el desempeño de una persona en una prueba con las inferencias que se puedan hacer de sus fortalezas y debilidades en el ámbito cognitivo (Leightony Gierl, 2007a; Nichols, 1994).

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 que se evidencien en la resolución de problemas con las interpretaciones que se puedan hacer sobre el desempeño. Tal modelo constituye un enfoque con el que se identifican y miden dichas habilidades y se explicará a continuación.

Modelos cognitivos para la evaluación cognitiva diagnóstica

Desde un enfoque cognitivo, se asume que la resolución de problemas necesita 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, Alves y Gotzmann, 2009).

En el campo de la medición educativa, 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 (Gierl, Roberts, Alves y Gotzmann, 2009; Leighton y Gierl, 2007a).

De esta manera, la evaluación cognitiva diagnóstica es una forma de evaluación que emplea un modelo cognitivo para desarrollar o identificar ítems que midan habilidades y conocimientos específicos, así como para llevar a cabo los análisis psicométricos de los patrones de respuesta de las personas para realizar inferencias a partir de ellos y brindar información relevante en términos de las debilidades y las fortalezas cognitivas que se evidencien (Gierl, Cui y Zhou, 2009).

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.

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, las habilidades de procesamiento y la manera en que son utilizadas por el estudiantado para producir la puntuación de una prueba.

Tradicionalmente, se utiliza la puntuación obtenida en una prueba para generar informes de resultados. No obstante, tal puntuación es un indicador que no permite detallar la forma en que el estudiantado piensa y resuelve situaciones educativas. De hecho, frecuentemente se asume que las personas que resuelven correctamente un problema han utilizado las habilidades y el conocimiento apropiados para tal efecto. Sin embargo, este supuesto puede ser falso, ya que se ha demostrado que las personas pueden obtener respuestas correctas utilizando conocimientos y habilidades que no se relacionan con el objetivo especificado en el

ítem (Norris, Leighton y Phillips, 2004) y de esta manera las inferencias que se pretendan hacer con respecto a las puntuaciones resultarían inadecuadas.

La evaluación cognitiva diagnóstica también se puede utilizar para vincular las teorías de cognición y aprendizaje con la enseñanza. En la actualidad, la mayoría de las pruebas en el ámbito educativo brindan poca información al estudiantado, docentes y padres y madres de familia acerca de por qué hay estudiantes que muestran un desempeño pobre o cuáles condiciones de la enseñanza se pueden modificar para el mejoramiento del aprendizaje (National Research Council, NRC, 2001). Por ello, se hace indispensable contar con información que señale algo más que el éxito o el fracaso y que permita decidir qué hacer ante la existencia de fallas en el aprendizaje. En ese sentido, la comprensión de las demandas cognitivas demandadas por las personas en los contenidos que aprenden, permitirá identificar los mecanismos que influyen en ellas para así construir ambientes de aprendizaje que las fomenten (Castañeda, 2004).

Atributos cognitivos: definición y jerarquías

La naturaleza de la realimentación que se puede obtener a partir de una evaluación cognitiva diagnóstica se basa, fundamentalmente, en la definición de las variables latentes en un contexto particular. Un *atributo cognitivo* –también conocido como *atributo*– es una descripción del conocimiento procedimental o declarativo que se necesita para llevar a cabo una tarea en un dominio específico (Birenbaum, Kelly y Tatsuoka, 1993; Gierl, Leighton, Wang, Zhou, Gokiert y Tan, 2009).

Por su parte, Leighton, Gierl y Hunka (2002) señalan que los atributos 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.

Los atributos se pueden definir con distintos grados de especificidad, según el objetivo de la evaluación cognitiva diagnóstica, la granularidad con que se quiera hacer afirmaciones acerca de las personas y la identificación de los procesos cognitivos subyacentes.

Por ejemplo, si se está investigando sobre restas con fracciones, las tareas son más delimitadas que las que se necesitan en un estudio sobre desempeño en operaciones aritméticas básicas (que incluyen las restas con fracciones). Por ello, la granularidad en la definición de atributos aumenta el alcance de la tarea que se analiza y refleja un aumento en la complejidad cognitiva de las acciones involucradas en la resolución (Rupp, Templin y Henson, 2010). No obstante, se debe tomar en cuenta que aunque es posible descomponer atributos

individuales en tareas más complejas, esto también aumenta la cantidad de atributos para los ítems, lo cual puede hacer imposible 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. Ante todo, la granularidad tiene que ver con el objetivo de la evaluación. A modo de ejemplo, en la tabla 1 se muestra una lista de atributos elaborada para el tema "resolución de ecuaciones algebraicas".

Tabla 1
Ejemplo de definición de atributos para un tema específico

Atributo	Descripción
A1	Comprensión del significado de símbolos y convenciones
A2	Comprensión de la descripción textual de un problema
A3	Efectuar manipulaciones algebraicas
A4	Resolución de ecuaciones lineales
A5	Resolución de ecuaciones cuadráticas
A6	Resolución de sistemas de ecuaciones
A7	Elaboración de una representación tabular
A8	Elaboración de una representación gráfica

Fuente: Traducción libre a partir de Gierl, Leighton y Hunka (2007).

Rupp, Templin y Henson (2010) señalan tres características distintivas para los atributos pertenecientes a los ítems de una prueba:

- 1) la etiqueta del atributo: se refiere a una palabra o frase que refleja el significado principal del atributo;

- 2) la definición del atributo: es un párrafo o texto corto que describe las diferentes facetas de un atributo con mayor detalle; y
- 3) las instrucciones para la codificación: son especificaciones que ayudan a los especialistas a determinar si un atributo pertenece o no a un ítem, es decir, si un ítem en particular lo mide o no.

Los atributos para una prueba pueden identificarse utilizando diferentes métodos (por ejemplo, opinión de expertos, análisis de tareas, respuestas escritas de estudiantes). En particular, las revisiones de los ítems y los análisis de protocolos sirven para estudiar los requerimientos de las tareas necesarias para su resolución. Dichas revisiones, por lo general, se llevan a cabo para identificar el conocimiento y las habilidades requeridas para resolver ítems según el criterio de especialistas (por ejemplo, quienes se encargan de elaborar pruebas) que están familiarizados con los contenidos del área que se está midiendo, cómo se confecciona una prueba y la forma en que el estudiantado resuelve problemas.

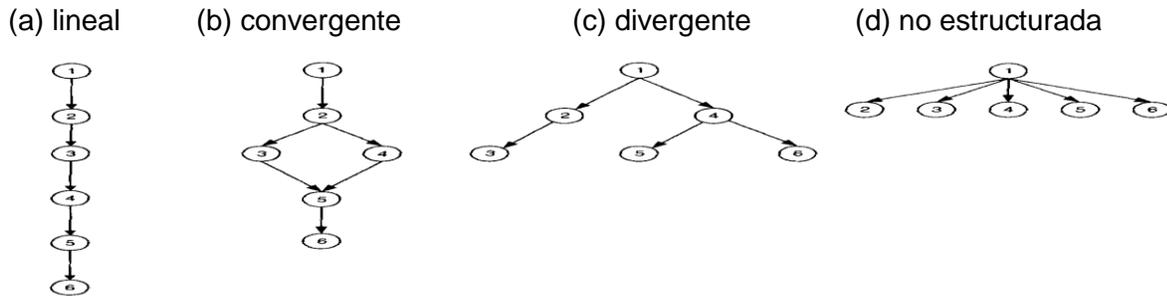
También se les puede preguntar a estudiantes que piensen en voz alta cómo resuelven problemas y los análisis de protocolos (verbales y escritos) se pueden utilizar para estudiar los procesos cognitivos involucrados en dicha resolución.

Tipos de jerarquías de atributos

Diversos estudios han permitido identificar jerarquías prototípicas de atributos que muestran las relaciones de dependencia que pueden existir entre ellos en un modelo de diagnóstico cognitivo. En la figura 1 se muestran algunas de estas jerarquías.

Figura 1

Jerarquías prototípicas de atributos



Fuente: Gierl, Leighton y Hunka (2007)

De manera análoga a la teoría de grafos, se puede hablar que esas estructuras jerárquicas son árboles de atributos que contienen ramas conformadas por cadenas de atributos. En una jerarquía lineal, todos los atributos están ordenados secuencialmente en una cadena simple, por lo que una persona que domine un atributo que esté al final de la cadena implica que también domina todos los atributos que le preceden. En la tabla 2 se muestran todos los posibles perfiles de atributos para las personas a partir de la jerarquía lineal presentada en la figura 1(a).

Tabla 2

Posibles perfiles para una jerarquía lineal de 6 atributos

	Atributo 1	Atributo 2	Atributo 3	Atributo 4	Atributo 5	Atributo 6
Perfil 1	0	0	0	0	0	0
Perfil 2	1	0	0	0	0	0
Perfil 3	1	1	0	0	0	0
Perfil 4	1	1	1	0	0	0
Perfil 5	1	1	1	1	0	0
Perfil 6	1	1	1	1	1	0
Perfil 7	1	1	1	1	1	1

Fuente: Rupp, Templin y Henson (2010)

Como se ve en la tabla 2, hay 7 posibles perfiles de atributos. Un 1 en una casilla determinada indica que se domina el atributo; en caso contrario, se escribe un 0.

Por su parte, en una jerarquía convergente (ver figura 1(b)), un atributo puede tener múltiples prerrequisitos en una cadena, por lo que si una persona domina uno de los atributos ubicados en la parte inferior de la cadena, entonces domina al menos uno de los atributos anteriores a él o quizás todos los que lo preceden.

En la figura 1(c) se muestra una jerarquía divergente. En dicha estructura hay múltiples ramas que se originan a partir de un mismo atributo. Por ejemplo, hay dos ramas que se originan del atributo 1. La rama que contiene los atributos 2 y 3 es una cadena lineal y se distingue de la que forman los atributos 4, 5 y 6 que es otra cadena divergente. Por lo tanto, si se sabe que una persona domina un atributo cualquiera de alguna de las ramas, entonces, también domina el primer atributo que es común a todas las ramas. Una jerarquía divergente representa rutas separadas hacia el dominio de distintos conjuntos de atributos (Rupp, Templin y Henson, 2010).

Finalmente, en una jerarquía no estructurada (figura 1(d)), un atributo es prerrequisito de varios atributos a la vez, por lo que si una persona domina cualquiera de los atributos (del 2 al 6, en el caso de la figura 1(d)), esto implica que domina el atributo 1, pero no necesariamente domina cualquiera de los otros atributos de la jerarquía.

Estas estructuras jerárquicas presentadas como prototipos de relaciones de dependencia entre atributos se pueden combinar para formar redes más complejas entre atributos, en las que la complejidad varía según la tarea cognitiva asociada a la resolución del problema particular (Gierl, Leighton y Hunka, 2007).

El método de representación del espacio de reglas

Tatsuoka (1983, 1990, 2009) propuso el método de representación del espacio de reglas (Artavia-Medrano y Larreamendy-Joerns, en prensa) denominado *rule space* (de aquí en adelante, método de representación del espacio de reglas, MRER), el cual es un enfoque probabilístico para brindar diagnósticos cognitivos a partir de análisis que se hacen acerca de si una persona evidencia poseer una lista de habilidades cognitivas que se necesitan para llevar a cabo una tarea (por ejemplo, resolver un problema o en particular, un ítem de una prueba), para lo cual combina la teoría de respuesta al ítem con información basada en habilidades cognitivas (Chen, 2006).

Los análisis propios del MRER se llevan a cabo en varias etapas. En la primera de ellas, se identifican las habilidades necesarias para resolver los ítems de una prueba, es decir, los atributos de tales ítems. Una vez que los atributos se han identificado, se elabora la llamada *matriz de incidencia* o matriz *Q*, en la cual se evidencian las relaciones entre los atributos y los ítems de la prueba: las columnas representan los ítems y las filas corresponden a los atributos,

por lo que cada elemento de la matriz Q indica si el ítem requiere o no de un atributo en particular (lo cual se denota con un 1 o un 0, respectivamente).

A modo de ilustración se presenta la siguiente matriz Q :

$$Q = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Como puede verse, la matriz anterior muestra la relación entre 3 ítems y 2 atributos. Además, el ítem 1 requiere de los atributos 1 y 2; el ítem 2 únicamente requiere del atributo 1 y el ítem 3 incluye solamente el atributo 2.

A partir de una matriz Q y de los patrones de respuesta de las personas, es posible inferir patrones de dominio de atributos. Se utiliza álgebra booleana para generar todas las posibles combinaciones entre los patrones de respuesta y los patrones de dominio de atributos.

Como el MRER asume que para que una persona acierte un ítem debe entonces dominar todos los atributos que dicho ítem involucra, en la tabla 3 se muestran los patrones de respuesta asociados a los patrones de dominio de atributos, para lo cual se tomará como referencia la matriz Q anteriormente descrita.

Tabla 3

Relación entre dominio de atributos y patrones de respuestas

Patrón de dominio de atributos		Patrón asociado de respuestas a los ítems		
Atributo 1	Atributo 2	Ítem 1	Ítem 2	Ítem 3
0	0	0	0	0
1	0	0	1	0
0	1	0	0	1
1	1	1	1	1

Fuente: Dogan y Tatsuoka (2008)

Los patrones de respuesta asociados a los posibles patrones de dominio de atributos son llamados *patrones ideales de respuesta* y, una vez que se ha creado la lista de tales patrones, el MRER determina cuáles de estos patrones están más cerca de las respuestas dadas por las personas. Para ello, se utiliza el criterio de la distancia de Mahalanobis (D^2). Una asociación aceptable entre los patrones ideales y las respuestas dadas a los ítems determina la

clasificación de examinados en *estados de conocimiento*. La probabilidad de que una persona en particular pertenezca a un posible estado de conocimiento se calcula en función de D^2 . Finalmente, las probabilidades de dominio de atributos para cada persona se calculan utilizando ponderaciones entre las probabilidades de distancia entre los patrones ideales y los patrones observados.

El método desarrollado por Tatsuoka supera las limitaciones del análisis tradicional de pruebas, pues ofrece información sobre las estrategias, los procesos y las representaciones de conocimiento que evidencian las personas (Dogan y Tatsuoka, 2008), por lo que permite analizar la efectividad de los ítems en la medición de constructos e identificar la precisión con que se miden los atributos involucrados, así como brindar información acerca de las fortalezas y las debilidades cognitivas del estudiantado.

El método de jerarquía de atributos

El enfoque propio del MRER indica que los atributos cognitivos no necesariamente comparten relaciones jerárquicas o de dependencia entre ellos (Tatsuoka, 1990). No obstante, un grupo de investigadores de la Universidad de Alberta, Canadá, plantearon en el año 2002, un método derivado del MRER –en tanto que utilizan sus conceptos fundamentales, tales como atributos cognitivos y matriz de incidencia, entre otros– en el cual se evidenciará que las habilidades cognitivas no funcionan aisladamente, sino que pertenecen a una red de procesos interrelacionados (Leighton, Gierl y Hunka, 2002, 2004).

El *attribute hierarchy method* (de aquí en adelante, método de jerarquía de atributos, MJA) está diseñado para vincular explícitamente la teoría cognitiva con la práctica psicométrica, específicamente en cuanto a facilitar el desarrollo y el análisis de pruebas psicológicas y educativas (Leighton, Gierl y Hunka, 2004). El método ilustra cómo el desempeño cognitivo puede ser evaluado utilizando el enfoque de procesamiento de la información, pues el MJA necesita de un modelo cognitivo de atributos estructurados para evaluar el desempeño de las personas. Tal estructura de atributos se conoce como *jerarquía de atributos* y define el orden psicológico entre los atributos que se necesitan para resolver un ítem de una prueba (Gierl, Leighton y Hunka, 2007).

El orden de los atributos se puede derivar a partir de consideraciones cognitivas propias de su estructura (es decir, cuáles atributos funcionan como prerrequisitos de otros), o bien, mediante consideraciones procedimentales (la forma en que se utilizan en la resolución de un

problema). En suma, la jerarquía de atributos representa tanto el constructo como el conocimiento y las habilidades de procesamiento que subyacen al desempeño en una prueba.

Componentes cognitivos del MJA

Como ya se indicó, el MJA se basa en el supuesto de que el desempeño en una prueba depende de un conjunto de procesos cognitivos o habilidades que se necesitan para resolver correctamente los ítems de dicha prueba y que están ordenados de manera jerárquica.

De acuerdo con Leighton, Gierl y Hunka (2004), los componentes cognitivos del MJA son: 1) la identificación de atributos cognitivos, 2) la especificación de la jerarquía de atributos para el modelo de desempeño en la prueba, y 3) la consideración de estructuras jerárquicas.

La identificación de atributos cognitivos es un aspecto medular del MJA, pues las inferencias de carácter cognitivo que se hagan del desempeño de una persona dependen de la precisión con que se haya identificado el orden psicológico y procedimental de las habilidades cognitivas que se requieren para la resolución de un ítem. Por ello, la jerarquía identificada resulta ser una hipótesis del desempeño cognitivo en el dominio específico de interés (Leighton, Gierl y Hunka, 2004).

De manera análoga al MRER, los atributos se identifican y estudian utilizando técnicas propias de la psicología cognitiva, tales como: análisis racional de tareas, análisis de protocolos verbales de estudiantes, y consulta a expertos en el dominio específico, entre otros (Buck, VanEssen, Tatsuoka, Kostin, Lutz y Phelps, 1998; Chen, 2006; Crandall, Klein, y Hoffman, 2006). Asimismo, las revisiones de ítems suelen llevarse a cabo por especialistas en diseño de pruebas y que estén familiarizados con el área de contenidos medida en la prueba, así como con las estrategias utilizadas por el estudiantado en la solución de problemas.

En el MJA, a diferencia del MRER, la relación de dependencia entre atributos debe identificarse previamente al desarrollo de ítems, para que se pueda tener un máximo control sobre los atributos específicos que cada ítem mide y lograr con ello crear una única matriz adyacente. Esta quizás sea la diferencia fundamental entre ambos métodos: mientras que en el MJA se asume que las pruebas son sensibles a la manera en que se organizan los atributos, en el MRER no se considera tal estructura, por lo que en este último es posible identificar atributos en pruebas ya existentes.

En cuanto al segundo componente cognitivo, es decir, la especificación de la jerarquía de atributos para el modelo de desempeño en la prueba, cabe resaltar que tal organización puede derivarse de consideraciones empíricas (por medio de análisis de protocolos, por ejemplo) o

consideraciones teóricas (serie de secuencias de etapas de desarrollo propuesta por Piaget: preoperacional, operacional concreta y operacional formal), por las que la jerarquía especificada resulta ser el modelo cognitivo para el desarrollo de pruebas y para explicar el desempeño de las personas (Leighton, Gierl y Hunka, 2004).

Finalmente, en lo que se refiere a la consideración de estructuras jerárquicas, tales especificaciones son las relaciones de dependencia entre atributos, por lo que representarán las hipótesis acerca de cuáles perfiles de atributos deberían observarse en una muestra de estudiantes. De acuerdo con Rupp, Templin y Henson (2010), el número de perfiles de atributos que se pueden identificar para una estructura jerárquica determinada, aumenta significativamente según la cantidad de atributos, especialmente si tales atributos son mayormente independientes entre sí –como en el caso de una jerarquía no estructurada–. No obstante, el número de perfiles de atributos se puede reducir considerablemente al postular relaciones de dependencia entre los atributos que serán medidos en una prueba.

Componentes psicométricos del MJA

Según Leighton, Gierl y Hunka (2004), los componentes psicométricos del MJA son: 1) la representación formal de la jerarquía, 2) la generación de patrones esperados de respuesta, 3) la estimación de probabilidades de respuesta a los ítems, y 4) la clasificación de patrones observados de respuesta.

Los componentes psicométricos anteriormente enumerados, y que se detallarán a continuación, requieren del desarrollo de un conjunto de matrices (adyacente, de accesibilidad y de incidencia) que fueron introducidas por Tatsuoka (1983, 1990, 2009) y que en el MJA se derivan de la jerarquía de atributos diseñada para facilitar la elaboración de los ítems de una prueba y la construcción de estados potenciales de conocimiento para las personas que, en este modelo, se expresan en términos de los perfiles de atributos si la jerarquía es verdadera.

Una vez que se ha identificado la jerarquía de atributos, esta se puede representar matemáticamente por una matriz binaria A , llamada *matriz adyacente* de orden $k \times k$, donde k es el número de atributos. En dicha matriz se especifica la relación directa que existe entre cada par de atributos. De esta manera, el elemento a_{ij} de la matriz A indica si el atributo i es un prerequisite directo del atributo j ; en caso afirmativo, se tiene $a_{ij} = 1$, o de lo contrario, $a_{ij} = 0$.

A manera de ejemplo, considérese la figura 1(c). En ella se muestra que el atributo 1 es un prerequisite directo para el atributo 2, el cual, a su vez, es prerequisite directo para los atributos 3 y 4. Además, el atributo 4 es prerequisite directo para los atributos 5 y 6. Dicha

jerarquía se representa en la matriz de incidencia A de orden 6×6 , en la que los elementos a_{12} , a_{23} , a_{24} , a_{45} y a_{46} son iguales a 1 (según lo anteriormente descrito).

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Para especificar las relaciones directas e indirectas entre atributos, se utiliza una matriz R , llamada *matriz de accesibilidad*, de orden $k \times k$, donde k es el número de atributos. La matriz R se calcula con $R = (A + I)^n$, donde n es el entero requerido para que R alcance la invarianza, $n = 1, 2, \dots, k$; A es la matriz adyacente, e I es la matriz identidad. R se puede transformar mediante un conjunto de sumas booleanas a las filas de la matriz adyacente. La j -ésima fila de la matriz R especifica todos los atributos para el cual el atributo j -ésimo es un prerequisite directo o indirecto de otros atributos y un prerequisite de sí mismo (esto último se puede notar en la existencia de unos en toda la diagonal principal de la matriz R).

Como ejemplo, se presenta la matriz de accesibilidad R , en la que se muestran las relaciones directas e indirectas entre seis atributos de la jerarquía mostrada en la figura 1(c), calculada a partir de la matriz A .

$$R = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

El conjunto de ítems que miden todas las posibles combinaciones de atributos cuando son independientes, es decir, cuando la correspondiente matriz R es la matriz identidad, se conoce como *conjunto potencial de ítems*. El tamaño del conjunto potencial es $2^k - 1$, donde k es el número de atributos. De esta manera, incluso para un valor pequeño de k , el conjunto potencial de ítems es muy grande. La representación matemática de este conjunto se hace en una

no corresponderá a la jerarquía de atributos y, por tanto, debe eliminarse. A continuación, se muestra la matriz Q_r propia de la situación que se ha venido describiendo.

$$Q_r = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Como se ve, la matriz reducida de incidencia es de orden 6×11 , lo que implica que de un conjunto potencial de 63 ítems, solo 11 corresponden lógicamente a la jerarquía de atributos diseñada previamente. La matriz Q_r es particularmente importante para la elaboración de pruebas, ya que puede utilizarse como las especificaciones cognitivas para su construcción (Leighton, Gierl y Hunka, 2004). Por ello, al describir los requerimientos cognitivos para un dominio específico mediante una jerarquía de atributos y al especificar los ítems necesarios para medirlos, el MJA posibilita la vinculación directa entre aspectos cognitivos de las personas y la elaboración de pruebas (Gierl, Leighton y Hunka, 2007).

El segundo componente psicométrico del MJA es la generación de *patrones esperados de respuesta*. Leighton, Gierl y Hunka (2004) emplean este término en vez del que utiliza Tatsuoka (1991, 1995, 2009) en sus planteamientos –patrones ideales de respuesta–, pues consideran que los patrones de respuestas deberían producirse solo si la jerarquía de atributos es verdadera, según se especifica en la matriz A .

Relacionados a los patrones esperados de respuesta están los *examinados "esperados"*, es decir, quienes utilizan los atributos de manera consistente con la jerarquía definida. Asimismo, se considera que tales personas no cometen errores aleatorios [slips] o errores que producen inconsistencias entre los patrones observados y esperados de respuestas. Utilizando operaciones booleanas se crea la *matriz de respuestas esperadas* o matriz E de orden $i \times j$, donde j es el número de personas e i es el número reducido de ítems que resulta de las restricciones impuestas por la jerarquía (Gierl, Zheng y Cui, 2008).

En la tabla 4 se muestran los patrones esperados de respuesta, es decir, aquellos que pueden ser claramente explicados por la presencia o la ausencia de atributos sin cometer errores aleatorios. Tales patrones se elaboraron según la jerarquía de la figura 1(c).

Tabla 4

Patrones esperados de respuesta para la jerarquía de la figura 1(c)

Patrón de atributos	Patrón esperado de respuestas	Puntuación total
100000	10000000000	1
110000	11000000000	2
111000	11100000000	3
110100	11010000000	3
111100	11111000000	5
110110	11010100000	4
111110	11111110000	7
110101	11010001000	4
111101	11111001100	7
110111	11010101010	6
111111	11111111111	11

Fuente: Cui (2007)

A manera de ejemplo, la cuarta fila de la tabla 4 se puede interpretar así: el patrón de atributos (110100) indica que una persona que domine los atributos 1, 2 y 4 debería producir el patrón esperado (11010000000) y obtener una puntuación total de 3, si la jerarquía es verdadera. Aquí se evidencia que la puntuación total en una prueba no brinda suficiente información acerca de las debilidades y las fortalezas de una persona, pues, por ejemplo, una puntuación total de 3 también se puede conseguir con el patrón de atributos (111000) que produce el patrón esperado (11100000000) y que refleja una información distinta a la anterior en términos cognitivos, pues en este último caso se afirma que la persona domina los atributos 1, 2 y 3 y que se espera que acierte los ítems 1, 2 y 3. En suma, dos personas con la misma puntuación total no necesariamente poseen los mismos patrones de atributos.

Los dos últimos componentes psicométricos del MJA se pueden basar en la utilización de procedimientos propios de la teoría de respuesta al ítem (TRI), o bien, otros que no se basan en la TRI y que se emplean para evaluar el ajuste de los datos al modelo y así calcular la probabilidad del dominio de atributos. En uno y otro caso se parte del supuesto de que las respuestas de las personas son consistentes con la jerarquía de atributos establecida; en el primero de ellos se calculan *curvas características esperadas* para cada ítem utilizando un modelo de teoría de respuesta al ítem (TRI). Una vez que se han estimado los parámetros del ítem y el valor de θ asociado a cada patrón esperado de respuesta, se calcula la probabilidad de acertar cada ítem según el modelo de TRI para cada patrón esperado. El detalle del procedimiento se encuentra en Leighton, Gierl y Hunka (2004).

La otra manera de abordar la clasificación de patrones observados de respuesta no utiliza procedimientos basados en la TRI. Al respecto Gierl, Cui y Hunka (2007) así como Gierl, Leighton y Hunka (2007) describen cómo se puede estimar la probabilidad de que una persona posea combinaciones específicas de atributos basado en su patrón de respuestas. Para ello utilizan redes neurales artificiales o modelos conexionistas, los cuales son "sistemas informáticos reticulares (de inspiración neuronal) que aprenden de la experiencia mediante la automodificación de sus conexiones" (Pitarque, Roy y Ruiz, 1998, p. 388). Por su parte, Gierl y Zhou (2008, p. 32) definen red neural como "un tipo de arquitectura de procesamiento paralelo que transforma un estímulo recibido por una unidad de entrada en una señal para una unidad de salida mediante una serie de unidades ocultas".

En el caso del MJA, las unidades de entrada son los patrones observados de respuestas de las personas a los ítems. Las unidades de salida son las probabilidades de que una persona posea un atributo ilustrado en la jerarquía (Cui, 2007).

Debido a la existencia de errores aleatorios [slips] que provocan incertidumbre en los valores de las probabilidades de dominio de atributos, los patrones observados de respuestas no son consistentes con los patrones esperados de respuestas (dado que estos últimos se infieren del modelo cognitivo). Tales errores pueden ocurrir por diferentes razones, entre otras: los atributos no fueron identificados con precisión, la jerarquía de atributos que se especificó es inapropiada, los ítems no miden los atributos de la jerarquía, la prueba resulta ser inadecuada para la muestra de estudiantes o bien, que estos producen respuestas de manera aleatoria (Gierl y Zhou, 2008).

De esta manera, la relación entre los vectores esperados de respuesta (a partir de la matriz E) y los vectores de atributos asociados a ellos se establece presentando a la red cada patrón esperado, lo cual se hace repetidamente hasta que el error de la red neural alcance un nivel aceptable. Una vez que dicha relación se ha establecido de manera satisfactoria, se produce un conjunto de matrices que permiten –por medio de la ponderación– transformar cualquier vector de respuestas observadas en el vector de atributos asociado a él y con ello se logra calcular las probabilidades de dominio de atributos.

La ventaja de utilizar redes neurales radica en ser un enfoque que no depende de modelos de TRI ni de los supuestos básicos acerca de las propiedades de los parámetros asociadas a su distribución. Asimismo, permiten estimar las probabilidades de que las personas dominen cada atributo minimizando el error asociado a tal estimación (Cui, 2007).

¿Cómo evaluar el ajuste de los datos al modelo?

Cuando se comparan los patrones esperados de respuesta –producidos en la matriz E – con los patrones observados de respuestas, pueden surgir algunas discrepancias, por lo que se hace necesario evaluar la consistencia entre ambos patrones. El estudio del ajuste de datos al modelo se puede llevar a cabo calculando el *índice de consistencia de la jerarquía* (de aquí en adelante ICJ), el cual evalúa el grado en que los patrones observados de un gran número de examinados resultan ser consistentes con los patrones esperados de respuesta que se generan a partir de la jerarquía de atributos (Cui y Leighton, 2009; Cui, Leighton, Gierl, y Hunka, 2006).

El dominio de atributos sucede cuando una persona acierta los ítems que involucran tales atributos. Por ello, el ICJ para la persona j se define como:

$$ICJ_j = 1 - \frac{2 \sum_{i \in S_{correct\ j}} \sum_{g \in S_i} X_{ji} (1 - X_{jg})}{N_{c_j}}$$

donde X_{ji} es la puntuación de la persona j (0 ó 1) para el ítem i , S incluye solo los ítems que miden los atributos comprendidos en la jerarquía para el ítem i y N_{c_j} es el número total de comparaciones con los ítems acertados por la persona j . Si dicha persona contesta correctamente el ítem i , $X_{ji} = 1$, entonces se espera que también acierte el ítem g que corresponde a S , $X_{jg} = 1$ ($g \in S_i$), donde S_i es el subconjunto de ítems que la persona j responde correctamente. Sin embargo, si $X_{jg} = 0$, entonces $X_{ji} (1 - X_{jg}) = 1$, lo cual es considerado como un desajuste del vector de respuestas j con respecto a la jerarquía y de ahí que la expresión $\sum_{i \in S_{correct\ j}} \sum_{g \in S_i} X_{ji} (1 - X_{jg})$ represente el número total de desajustes.

El valor del ICJ tiene un rango de -1 a 1 . Para efectos de su interpretación, Cui (2007) ofrece los siguientes criterios:

- si $ICJ = 1$, entonces el vector de respuestas de una persona se ajusta perfectamente a la jerarquía de atributos, es decir, las respuestas de la persona corresponden a uno de los patrones esperados de respuesta;
- si $ICJ = -1$, entonces el vector de respuestas se desajusta completamente de la matriz Q_n , es decir, la persona acierta un ítem, pero falla todos los demás ítems que requieren

del conjunto de atributos medidos en ellos y que se necesitaban también para el ítem que sí acertó;

- c) si los valores de ICJ se acercan a -1 , se puede concluir que la persona posiblemente utiliza, en la resolución del ítem, habilidades y conocimientos diferentes a los especificados en la jerarquía de atributos y en la correspondiente matriz Q_r y en este caso el modelo no se puede utilizar para hacer inferencias cognitivas válidas acerca del desempeño de las personas; y
- d) si los valores de ICJ se acercan a 1 , esto sugiere un mejor ajuste de las respuestas de las personas de conformidad con la jerarquía establecida.

Asimismo, Cui (2007) advierte que dependiendo de la forma de la distribución del ICJ, se pueden utilizar la media aritmética o la mediana como indicadores de ajuste general de los datos al modelo. En este caso, un valor alto de tales estadísticos se podría interpretar como un ajuste general de las respuestas de estudiantes a la jerarquía de atributos especificada.

Una forma de estimación de la confiabilidad de los atributos

La confiabilidad de los atributos se refiere a la precisión de las decisiones basadas en las puntuaciones de una prueba –generalmente con propósitos diagnósticos– acerca del dominio que tienen las personas de los atributos correspondientes (Gierl, Cui y Zhou, 2009). Un procedimiento para estimar la confiabilidad de un atributo es calcular la razón entre la varianza de la puntuación verdadera y la varianza de la puntuación observada en los ítems que miden cada atributo.

Dado que según el MJA cada ítem está destinado a medir una combinación de atributos en una jerarquía específica, se puede afirmar que cada atributo contribuye a una parte de la varianza a nivel del ítem. Por ello, el índice incorpora el concepto de dependencia de atributos en el cálculo de la confiabilidad utilizando el concepto de consistencia interna.

Para aislar la contribución de cada atributo al desempeño de cada persona en los ítems, la puntuación debe ponderarse según la diferencia de dos probabilidades condicionales. La primera probabilidad se asocia con el dominio de los atributos y la segunda con no dominarlos, es decir, una persona que domina el atributo puede acertar el ítem, pero una que no lo domina no puede contestarlo correctamente.

Gierl, Cui y Zhou (2009) adaptaron la fórmula del coeficiente alfa de Cronbach para estimar la confiabilidad de los atributos según los procedimientos del MJA y está dada por:

$$\alpha_{MJA_k} = \frac{n_k}{n_k - 1} \left[1 - \frac{\sum_{i \in S_k} W_{ik}^2 \sigma_{X_i}^2}{\sigma^2 \sum_{i \in S_k} W_{ik} X_i} \right]$$

en la que α_{MJA_k} es la confiabilidad para el atributo k , n_k es el número de ítems que miden el atributo k según la matriz Q_r (es decir, el número de elementos en S_k), $\sigma_{X_i}^2$ es la varianza de las puntuaciones observadas para el ítem i , $\sum_{i \in S_k} W_{ik}^2 \sigma_{X_i}^2$ es la suma de las varianzas ponderadas de las puntuaciones observadas en los ítems que miden el atributo k , y $\sigma^2 \sum_{i \in S_k} W_{ik} X_i$ es la varianza de las ponderaciones de las puntuaciones totales observadas dadas por $\sum_{i \in S_k} W_{ik}^2 \sigma_{X_i}^2$.

Dado que en el MJA, la etapa de reconocimiento estadístico de patrones se utiliza para estimar la probabilidad de que una persona posea una combinación específica de atributos, el cálculo de la confiabilidad de atributos puede utilizarse para mejorar la elaboración de informes de resultados con carácter diagnóstico, pues permite la creación de intervalos de confianza alrededor de las puntuaciones basadas en el dominio de atributos y de esta manera, poder crear perfiles con información específica acerca de las fortalezas y las debilidades cognitivas de las personas.

Consideraciones finales

En el presente artículo se ha presentado una revisión de una de las propuestas que más se ha desarrollado en los últimos ocho años en el campo de la evaluación cognitiva diagnóstica: el método de jerarquía de atributos.

Su necesidad se ha evidenciado por los requerimientos que cada vez aumentan en torno a disponer de modelos que permitan interpretar las puntuaciones de las personas de una manera diferente a la que tradicionalmente se ha hecho, pues se recurre a las habilidades cognitivas que subyacen a su pensamiento. De esta manera, a diferencia de los enfoques psicométricos tradicionales, es posible ahora contar con modelos cognitivos que sirvan como guía para los procesos de enseñanza, pues se dispone de un conjunto de conocimientos y

habilidades interconectados que explican, en alguna medida, la psicología del desempeño en la resolución de situaciones de un dominio específico.

El desarrollo de métodos que permitan vincular teorías cognitivas y del aprendizaje con prácticas psicométricas y de enseñanza es un campo promisorio en la evaluación educativa, pues permite a quienes investigan hacer representaciones más complejas a nivel de relaciones jerárquicas, por lo que es posible también evaluar una mayor cantidad de modelos cognitivos para el desarrollo de ítems y pruebas. Consecuentemente, es factible tener un conocimiento más profundo sobre las habilidades cognitivas evidenciadas en el aprendizaje académico y una exploración más confiable mediante la identificación y la comprensión de los componentes que generan fallas en el aprendizaje, más que solo una descripción estadística de lo que logró o no una persona en particular. Ello posibilita planear estrategias didácticas en función de tales logros y también poder proponer acciones para el mejoramiento de las debilidades detectadas en las personas.

Dado lo novedoso de los enfoques de evaluación orientados a aspectos cognitivos en la literatura técnica, se necesitan más estudios empíricos en un dominio de conocimientos específico para poder diseñar modelos cognitivos que guíen la elaboración de ítems y la interpretación de las puntuaciones de las personas en una prueba, con lo cual las inferencias acerca de su desempeño tengan mayor sentido y fundamento en las prácticas psicológicas y educativas contemporáneas.

Referencias

- Artavia-Medrano, Álvaro y Larreamendy-Joerns, Jorge. (en prensa). **Información cognitiva a partir de pruebas de gran escala: el método de representación del espacio de reglas.**
- Bejar, Isaac. (1984). Educational Diagnostic Assessment. **Journal of Educational Measurement**, 21(2), 175-189.
- Bernad, Juan. (2000). **Modelo cognitivo de evaluación cognitiva.** Madrid: Narcea S.A.
- Birenbaum, Menucha, Kelly, Anthony & Tatsuoka, Kikumi. (1993). **Toward a Stable Diagnostic Representation of Students' Error in Algebra** (Technical Report RR-92-58-ONR). Nueva Jersey: Educational Testing Service.
- Buck, Gary, VanEssen, Tom, Tatsuoka, Kikumi, Kostin, Irene, Lutz, Donna & Phelps, Matthew. (1998). **Development, selection and validation of a set of cognitive and linguistic attributes for the SAT I Verbal: sentence completion section** (Research Report RR-98-23). Nueva Jersey: Educational Testing Service.
- Castañeda, Sandra. (2004). **Educación, aprendizaje y cognición: teoría en la práctica.** México: Editorial El Manual Moderno S.A.
- Chen, Yi-Hsin. (2006). **Cognitively diagnostic examination of Taiwanese mathematics achievement on TIMSS-1999.** Dissertation doctoral, Arizona State University.
- Crandall, Beth, Klein, Gary & Hoffman, Robert. (2006). **Working minds. A practitioner's guide to cognitive task analysis.** Cambridge, MA: MIT Press.
- Cui, Ying. (2007). **The Hierarchy Consistency Index: A Person-fit Statistic for the Attribute Hierarchy Method.** Dissertation doctoral, University of Alberta.
- Cui, Ying y Leighton, Jacqueline. (2009). The Hierarchy Consistency Index: Evaluating Person Fit for Cognitive Diagnostic Assessment. **Journal of Educational Measurement**, 46(4), 429-449.
- Cui, Ying, Leighton, Jacqueline, Gierl, Mark y Hunka, Steve. (2006, abril). **A person-fit statistic for the attribute hierarchy method: The hierarchy consistency index.** Documento presentado en la reunión del National Council on Measurement in Education, San Francisco, CA, EE. UU. Recuperado el 9 de setiembre del 2007, de <http://www.education.ualberta.ca/educ/psych/crame>
- Dogan, Enis y Tatsuoka, Kikumi. (2008). An international comparison using a diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R. **Educ Stud Math**, 68, 263-272.
- Ferrer, Guillermo. (2006). **Sistemas de evaluación de aprendizajes en América Latina: balance y desafíos.** Santiago: Alfabetas Artes Gráficas.

- Frederiksen, Norman. (1990). Introduction. En N. Frederiksen, R. Glaser, A. Lesgold y M. Shafto. (Eds.) **Diagnostic monitoring of skills and knowledge acquisition** (pp. ix-xvii). Nueva Jersey: Erlbaum.
- Gierl, Mark, Cui, Ying y Hunka, Steve. (2007, abril). **Using Connectionist Models to Evaluate Examinees' Response Patterns on Tests: An Application of the Attribute Hierarchy Method to Assessment Engineering**. Artículo presentado en Annual Meeting of the National Council on Measurement in Education (NCME), Chicago, IL.
- Gierl, Mark, Cui, Ying y Zhou, Jiawen. (2009). Reliability and Attribute-Based Scoring in Cognitive Diagnostic Assessment. **Journal of Educational Measurement**, **46**(3), 293-313.
- Gierl, Mark, Leighton, Jacqueline y Hunka, Steve. (2000). Exploring the Logic of Tatsuoka's Rule-Space Model for Test Development and Analysis. **Educational Measurement: Issues and Practice** **19**(3), 34-44.
- Gierl, Mark, Leighton, Jacqueline y Hunka, Steve. (2007). Using the Attribute Hierarchy Method to Make Diagnostic Inferences. En Jacqueline Leighton y Mark Gierl (Eds.) **Cognitive Diagnostic Assessment for Education** (pp. 242-274). Nueva York: Cambridge University Press.
- Gierl, Mark, Leighton, Jacqueline, Wang, Changjiang, Zhou, Jiawen, Gokiart, Rebecca y Tan, Adele. (2009). **Validating Cognitive Models of Task Performance in Algebra on the SAT**. Nueva York: The College Board.
- Gierl, Mark, Roberts, Mary, Alves, Cecilia y Gotzmann, Andrea. (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.
- Gierl, Mark, Zheng, Yinggan y Cui, Ying. (2008). Using the Attribute Hierarchy Method to Identify and Interpret Cognitive Skills that Produce Group Differences. **Journal of Educational Measurement**, **45**(1), 65-89.
- Gierl, Mark y Zhou, Jiawen. (2008). Computer Adaptive-Attribute Testing. A New Approach to Cognitive Diagnostic Assessment. **Zeitschrift für Psychologie**, **216**(1), 29-39.
- Gorin, Joanna. (2006). Test design with cognition in mind. **Educational Measurement: Issues and Practice**, **25**(4), 21-35.
- Huff, Kristen y Goodman, Dean. (2007). The Demand for Cognitive Diagnostic Assessment. En Jacqueline Leighton y Mark Gierl (Eds.) **Cognitive Diagnostic Assessment for Education** (pp. 19-60). Nueva York: Cambridge University Press.
- Junker, Brian y Sijtsma, Klaas. (2001). Cognitive Assessment Models with Few Assumptions, and Connections with Nonparametric IRT. **Applied Psychological Measurement**, **25**(3), 211-220.

- Leighton, Jacqueline y Gierl, Mark. (2007a). 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, Jacqueline y Gierl, Mark. (2007b). Why Cognitive Diagnostic Assessment? En Jacqueline Leighton y Mark Gierl (Eds.) **Cognitive Diagnostic Assessment for Education** (pp. 3-18). Nueva York: Cambridge University Press.
- Leighton, Jacqueline, Gierl, Mark y Hunka, Steve. (2002, abril). **The attribute hierarchy model for cognitive assessment**. Documento presentado en la reunión del National Council on Measurement in Education, Nueva Orleans, EE. UU. Recuperado el 9 de setiembre del 2007, de <http://www.education.ualberta.ca/educ/psych/crame>
- Leighton, Jacqueline, Gierl, Mark y Hunka, Steve. (2004). The Attribute Hierarchy Method for Cognitive Assessment: A Variation on Tatsuoka's Rule-Space Approach. **Journal of Educational Measurement**, **41**(3), 205-237.
- Messick, Samuel. (1984). The Psychology of Educational Measurement. **Journal of Educational Measurement**, **21**(3), 215-237.
- Mislevy, Robert, Steinberg, Linda y Almond, Russell. (2003). On the Structure of Educational Assessments. **Measurement: Interdisciplinary Research and Perspectives**, **1**(1), 3-62.
- National Research Council. (2001). **Knowing what students know: the science and design of educational assessment**. Washington: National Academy Press.
- Nichols, Paul. (1994). A framework for developing cognitively diagnostic assessments. **Review of Educational Research**, **64**, 575-603.
- Norris, Stephen, Leighto, Jacqueline y Phillips, Linda. (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.
- Pellegrino, James, Baxter, Gail y Glaser, Robert. (1999). Addressing the "Two Disciplines" Problem: Linking Theories of Cognition and Learning with Assessment and Instructional Practice. **Review of Research in Education**, **24**, 307-353.
- Pitarque, Alfonso, Roy, Juan Francisco y Ruiz, Juan Carlos. (1998). Redes neurales vs modelos estadísticos: simulaciones sobre tareas de predicción y clasificación. **Psicológica**, **19**, 387-400. Recuperado el 11 de mayo del 2010, de <http://www.uv.es/revispsi/articulos3.98/pitarque.pdf>
- Rupp, André. (2007). The Answer is in the Question: A Guide for Describing and Investigating the Conceptual Foundations and Statistical Properties of Cognitive Psychometric Models. **International Journal of Testing**, **7**(2), 1-31.

- Rupp, André, Templin, Jonathan y Henson, Robert. (2010). **Diagnostic Measurement: Theory, Methods and Applications**. Nueva York: The Guilford Press.
- Tatsuoka, Kikumi. (1983). Rule Space: An Approach for Dealing with Misconceptions Based on Item Response Theory. **Journal of Educational Measurement**, 20(4), 345-354.
- Tatsuoka, Kikumi. (1990). Toward an integration of item-response theory and cognitive error diagnosis. En Frederiksen, N., Glaser, R. Lesgold, A. & Shafto, M. (Eds.) **Diagnostic monitoring of skills and knowledge acquisition** (p. 453-488). Nueva Jersey: Erlbaum.
- Tatsuoka, Kikumi. (1991). **Boolean algebra applied to determination of universal set of misconception states**. Research Report RR-91-44. Nueva Jersey: Educational Testing Service.
- Tatsuoka, Kikumi. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. En Nichols, P., Chipman, S. & Brennan, R. (Eds.) **Cognitively diagnostic assessment** (p. 327-359). Nueva Jersey: Erlbaum.
- Tatsuoka, Kikumi. (2009). **Cognitive Assessment: An Introduction to the Rule Space Method**. Nueva York: Routledge Taylor & Francis Group.