

Información cognitiva a partir de pruebas de gran escala: el método de representación del espacio de reglas

Cognitive information based on large-scale tests: Representation's method of rule space

Recibido: junio 18 de 2010 | Revisado: enero 23 de 2011 | Aceptado: julio 13 de 2011

ÁLVARO ARTAVIA MEDRANO*

Universidad de Costa Rica, San José, Costa Rica

JORGE LARREAMENDY-JOERNS**

Universidad de los Andes, Bogotá, Colombia

RESUMEN

En los últimos años, ha existido una creciente demanda para que las pruebas estandarizadas de rendimiento académico brinden información de mayor utilidad para los procesos educativos, manteniendo las ventajas de su proceso técnico de elaboración, y agregando información diagnóstica cuyo objetivo sea mejorar las áreas evidenciadas como deficientes por los estudiantes. El presente artículo aboga por el método de representación del espacio de reglas (*rule space*) de Tatsuoka (1983, 2009) como una opción para incorporar los beneficios de la evaluación cognitiva diagnóstica a las prácticas actuales de evaluación. El artículo explica brevemente el proceso de resolución de problemas, según la teoría del procesamiento de la información; presenta las bases para el análisis de los ítems, con miras a la aplicación del método y lo expone con especial énfasis en la elaboración de la matriz *Q*, como modelo cognitivo que permite 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 los examinados.

Palabras clave autores

Evaluación cognitiva diagnóstica, método de representación del espacio de reglas, modelos cognitivos.

Palabras clave descriptores

Resolución de problemas, psicología cognitiva, evaluación psicológica.

ABSTRACT

Over the last years, there has been a growing demand for standardized achievement tests that provide useful information to educational processes, by adding diagnostic information to improve the students' deficient areas, while maintaining advantages in technical development. This paper advocates for Tatsuoka's *rule space* method (1983, 2009) as an option to incorporate the benefits of cognitive diagnostic assessment to current educational practices. The article briefly explains the process of problem solving according to some information processing theories, provides the basis for item analysis, and describes the rule space method with particular emphasis on the development of the *Q* matrix as a cognitive model that allows the interpretation of results. The aim of this method is to link performance on a test with specific inferences about the knowledge and skills of examinees.

Key words authors

Cognitive diagnostic assessment, rule space method, cognitive models.

Key words plus

Problem solving, cognitive psychology, psychological assessment.

SICI: 1697-9267(201206)11:2<599:ICPPGE>2.0.TX;2-Z

Para citar este artículo. Artavia, A. & Larreamendy-Joerns, J. (2012). Información cognitiva a partir de pruebas de gran escala: el método de representación del espacio de reglas. *Universitas Psychologica*, 11(2), 599-610.

* alvartavia@gmail.com

** jlarream@uniandes.edu.co

Introducción

Usualmente, las evaluaciones psicométricas de gran escala en el área de medición del conocimiento dan como resultado puntuaciones que reflejan el nivel de desempeño global de los examinados en un área de contenidos, pero no proveen información sobre los procesos cognitivos implicados en la resolución de los ítems o sobre las concepciones erróneas evidenciadas por los estudiantes. Las pruebas de gran escala dan cuenta preferiblemente de lo que una persona ha aprendido, pero no de lo que está en capacidad de aprender (Embretson, 1990). Dado el actual énfasis en competencias y en aprender a aprender, sería deseable que dichas pruebas dieran información sobre los procesos cognitivos responsables del aprendizaje en un dominio específico.

La elaboración de evaluaciones de gran escala con propósitos de evaluación cognitiva es un campo de reciente desarrollo en la literatura psicológica y educativa (Gierl, Zhou & Alves, 2008). Puesto que el diseño de pruebas con un propósito exclusivo (como, por ejemplo, la evaluación cognitiva diagnóstica) puede ser prohibitivamente costoso, cabe preguntarse cómo incorporar los beneficios de la evaluación diagnóstica a las prácticas de evaluación y de qué manera puede obtenerse información relevante sobre las estructuras de conocimiento y los procesos cognitivos que subyacen al rendimiento en una prueba determinada.

Tatsuoka (1983, 1990, 2009) ha propuesto el método denominado *rule space* (en adelante, método de representación del espacio de reglas [MRER]). El MRER es un enfoque probabilístico para el análisis de datos de pruebas que combina la teoría de respuesta al ítem con información basada en habilidades cognitivas (Chen, 2006). El método desarrollado por Tatsuoka supera las limitaciones del análisis tradicional de pruebas de gran escala, al ofrecer información sobre las estrategias, los procesos y las representaciones de conocimiento que evidencian los examinados (Gierl, Leighton & Hunka, 2000). El MRER permite dar retroalimentación individual a los examinados a partir de la descomposición de la multidimensionalidad del espacio latente en habilidades cognitivas o atribu-

tos de granularidad más fina. Asimismo, el método permite analizar la efectividad de los ítems en la medición de constructos e identificar la precisión con que se miden los atributos involucrados (Yan, Almond & Mislevy, 2004).

Dada la novedad del MRER en la literatura técnica, el propósito del presente artículo es describirlo como un método para generar información acerca del conjunto de habilidades cognitivas específicas implicadas en la resolución de ítems. El artículo está organizado en cuatro secciones. En la primera, se presenta el modelo cognitivo de resolución de problemas que subyace al MRER. En la segunda sección, se presentan las bases para el análisis de los ítems con miras a la aplicación del MRER. En la tercera sección, se expone el MRER con especial énfasis en la elaboración de la matriz Q , como modelo cognitivo que permite la interpretación de resultados. El artículo concluye con una sección de consideraciones generales.

Modelo de resolución de problemas subyacente al MRER

El MRER provee información sobre las estrategias, los procesos y las representaciones cognitivas que subyacen y explican el desempeño de los examinados en una prueba de conocimiento. Para esto, asume que los ítems de una prueba son problemas en el sentido cognitivo del término. En particular, el MRER adopta el modelo de resolución de problemas (y, por tanto, de ítems psicométricos) propuesto Newell y Simon (1972), cuyos presupuestos continúan vigentes luego de décadas de su formulación inicial (Holyoak, 1995; VanLehn, 1989).

Según Newell y Simon (1972), un problema es cualquier situación en la cual el solucionador se encuentra en una situación inicial y desea alcanzar una situación final, pero las acciones necesarias para avanzar de la situación inicial a la final no son obvias. La solución de un problema puede entenderse como una secuencia de situaciones ($E_i, E_{ii}, \dots, E_f$), en la que E_i representa las instrucciones del problema o situación inicial y E_f representa la meta o solución del problema. Para avanzar entre las situaciones del problema, el solucionador cuenta con *acciones* u

operadores que le permiten transformar una situación E_n en una situación E_{n+1} . Por ejemplo, en un problema de aritmética básica, una acción puede ser identificar los sumandos a partir del enunciado y otra acción puede ser obtener el resultado de la suma mediante la recuperación de hechos aritméticos en memoria a largo plazo. Cuando varias acciones se encadenan para el logro de una meta, las acciones encadenadas constituyen *estrategias*. Las estrategias y las acciones pueden enunciarse en términos de reglas de la forma SI-ENTONCES o reglas de producción.

El análisis del desempeño de una persona al resolver un problema distingue dos niveles: el *ambiente de la tarea* y el *espacio del problema* (Crandall, Klein & Hoffman, 2006; Newell & Simon, 1972). El ambiente de la tarea se refiere a la descripción objetiva de una situación problema, en el sentido de ser una descripción generada por alguien que dispone de las condiciones óptimas de procesamiento para su solución. Cuando se especifica el ambiente de la tarea deben identificarse: a) la situación inicial; b) la situación final; c) las acciones posibles; d) todas las posibles estrategias que conducen a la solución

del problema; e) la información que está al alcance del solucionador; f) las entidades (e.g., objetos, representaciones) que hacen parte de la situación problema y g) el conocimiento relevante.

Sin embargo, el análisis del ambiente de la tarea solo da una visión parcial de las características de la actividad. Es necesario llevar a cabo un análisis del espacio del problema; es decir, de la *representación* que el solucionador tiene de la tarea misma y que determina la manera como los individuos efectivamente transforman la situación problema por medio de sus acciones. El espacio del problema varía en función del conocimiento previo del sujeto, de los rasgos engañosos y salientes de la tarea, de las estrategias de solución anteriormente aprendidas o dominadas y de otras variables relativas al conocimiento y procesamiento de información.

En la Tabla 1 se muestra una lista de acciones, estrategias o, más generalmente, piezas de conocimiento implicadas en algunos de los ítems de matemática de la prueba TIMSS-R de 1999 (Chen, Gorin, Thompson & Tatsuoka, 2006; Tatsuoka, 2009). En la Tabla 2 se muestran las acciones, estrategias y piezas de conocimiento implicadas en

TABLA 1

Lista de atributos para explicar el desempeño en los ítems de matemática de la prueba TIMMS-R 99

Tipo	Código	Descripción
Contenido	C1	Conceptos básicos y operaciones con números enteros
	C2	Conceptos básicos y operaciones con fracciones y decimales
	C3	Conceptos básicos y operaciones de álgebra elemental
	C4	Conceptos básicos y operaciones de geometría en dos dimensiones
Procesos	P1	Traducción, formulación y entendimiento de ecuaciones y expresiones para resolver un problema
	P2	Conocimiento de aplicaciones de cálculo en aritmética y geometría
	P3	Conocimiento de aplicaciones de decisión en aritmética y geometría
	P4	Aplicación de reglas en álgebra y resolución de ecuaciones
	P5	Razonamiento lógico (incluye razonamiento de casos, habilidades de pensamiento deductivo, si... entonces, condiciones necesarias y suficientes y habilidades de generalización)
Habilidades	H1	Aplicación de propiedades y relaciones con números
	H2	Utilización de figuras, tablas, cuadros y gráficas
	H3	Utilización de figuras geométricas
	H4	Aproximación y estimación
	H5	Evaluación, verificación y revisión de opciones
	H6	Establecimiento de patrones y relaciones (habilidades de pensamiento inductivo)

Fuente: adaptación de Chen et al. (2006) y Tatsuoka (2009).

TABLA 2
Ejemplo de codificación de un ítem

Enunciado del ítem	Un remolque de esquí acuático tiene la forma de un triángulo isósceles. Si uno de los ángulos congruentes mide 65° , ¿cuál es la medida de x ?
Codificación	[H3] Se da una figura geométrica. [C4] Es un problema de geometría. [P3] Se debe aplicar conocimiento sobre las relaciones entre ángulos para obtener la solución (es decir, como la suma de las medidas de los ángulos de un triángulo es 180°, la medida del tercer ángulo es $180^\circ - (65^\circ + 65^\circ) = 50^\circ$). [P5] El valor de x se puede obtener restando 180° de la mitad de la medida del tercer ángulo. [P2] Esto es, $180^\circ - 25^\circ = 155^\circ$.

Fuente: adaptación de Tatsuoka (2009).

un ítem particular, identificadas a partir de la lista de atributos de la Tabla 1.

Un aspecto esencial es definir el nivel de granularidad o precisión con que se describen las acciones, estrategias o piezas de conocimiento implicadas en un problema. Las estrategias, acciones o elementos de conocimiento pueden describirse con granularidad fina. Por ejemplo, si un ítem de física incluye la solución de una ecuación lineal, una regla relevante podría ser:

SI [la meta es encontrar el valor de una variable en una ecuación]

ENTONCES [mover (constantes a la derecha de la ecuación) & (variables a la izquierda)].

En este caso, la regla se refiere a un aspecto específico del proceso de solución de la ecuación y, por tanto, es una regla altamente particular, lo cual quiere decir que es aplicable a la resolución de cualquier ecuación, pero no de cualquier problema de física. Sin embargo, uno podría proponer reglas o atributos más genéricos, como es el caso del atributo: “Habilidad para resolver procedimientos algebraicos”, en cuyo caso la regla se aplica no solo a ecuaciones, sino también a cualquier procedimiento algebraico. Las dos descripciones son equivalentes, aunque ofrecen información de mayor o menor especificidad.

La decisión en favor de descripciones de distinta granularidad depende de los propósitos del análisis y de la posibilidad de generar atributos que sean

suficientemente informativos pero suficientemente manejables, para llegar a conclusiones razonables a nivel global (Lohman, 2000). Como se puede ver en las Tablas 1 y 2, la granularidad de las acciones o estrategias identificadas para un ítem particular puede variar. El MRER implica un análisis de ítems en los términos mencionados.

Análisis de ítems en atributos cognitivos

Tatsuoka (1990) introdujo el término *atributo* para referirse a reglas de producción, operaciones procedimentales, subtareas cognitivas, piezas de conocimiento o aspectos de la tarea que se identifican para los ítems de una prueba. Leighton, Gierl y Hunka (2002, p. 4) señalan que los atributos son “una descripción del conocimiento procedimental o declarativo necesario para llevar a cabo una tarea en un dominio específico”. En tal sentido, los atributos se pueden ver como fuentes de complejidad cognitiva en el desempeño de una prueba. Los atributos no son invariables, sino que están ligados al espacio del problema, es decir, varían en función de la competencia de los estudiantes (Ohlsson, 2007).

La primera etapa del MRER es la identificación de las habilidades requeridas para resolver los ítems de una prueba. Esta etapa implica un análisis de tareas en el sentido anteriormente descrito. Esta fase inicia con la elaboración de una lista preliminar de atributos. Para ello, el analista puede recurrir a literatura previa, al análisis racional de tareas, al análisis de protocolos verbales de estudiantes y a la

consulta a expertos en el dominio específico, entre otros (Brown & VanLehn, 1982; Buck et al., 1998; Chen, 2006; Crandall et al., 2006).

Posteriormente, la lista inicial se analiza con el fin de depurarla para que incluya los atributos requeridos para los ítems de una prueba en particular. Un grupo de especialistas codifica los ítems, en términos de la presencia o ausencia de atributos requeridos en su resolución. Tal codificación produce una matriz cuyas filas representan los ítems y cuyas columnas representan los atributos. Dicha matriz se denomina *matriz de incidencia*. Asimismo, se elabora un manual de codificación para los ítems en el que se explicitan los atributos considerados como relevantes; en el manual cada atributo es ilustrado a través de un ítem (Dogan, 2006).

Las relaciones directas entre atributos se especifican en una matriz binaria A , llamada *matriz adyacente* de orden $k \times k$, donde k es el número de atributos. En una matriz adyacente, un 1 en la posición (j, k) indica que el atributo j está directamente asociado al atributo k en forma de prerrequisito (j precede a k); en caso contrario, se escribe un 0. Para ilustrar este concepto, se presenta la matriz A en la que se pueden observar las relaciones directas entre los seis atributos ($A_1, A_2, \dots, A_6$) que la conforman:

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

En la matriz anterior, la primera fila indica que el atributo A_1 es prerrequisito para los atributos A_2 y A_4 (representados en dicha fila por unos en las columnas segunda y cuarta). Una fila de ceros, como en la fila 3, indica que el atributo A_3 no es prerrequisito para ninguno de los otros atributos. Los unos que aparecen en la parte triangular superior de la matriz indican una relación direccional para los prerrequisitos (por ejemplo, A_1 es prerrequisito para A_2 y A_4 , pero A_2 y A_4 no son prerrequisitos

para A_1). Esta jerarquía entre atributos se muestra en la Figura 1.

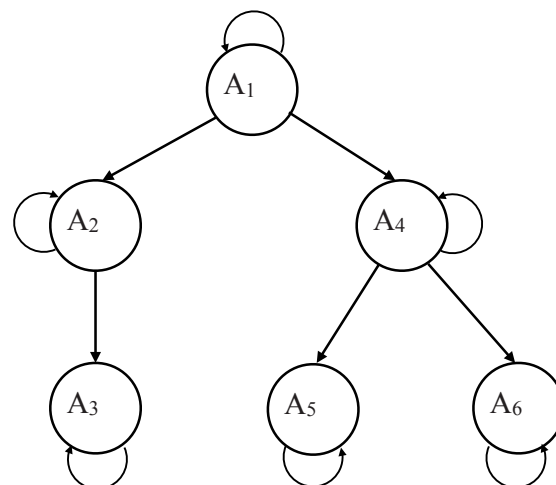


Figura 1. Relaciones directas e indirectas entre los atributos $A_1, A_2, \dots, A_6$

Fuente: Elaboración propia.

Nótese en la Figura 1 que el atributo A_1 representa las habilidades iniciales que se requieren que posea un examinado como prerrequisito para los demás atributos. Asimismo, la figura representa relaciones de prerrequisito directas (por ejemplo, A_2 en relación con A_3) e indirectas (por ejemplo, A_1 en relación con A_6). La relación entre atributos es independiente del conjunto de ítems que se utilice (Tatsuoka, 2009). 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. Esta matriz especifica todas las formas posibles de relacionar los atributos entre sí, por lo que provee información acerca de la estructura cognitiva de los atributos (Gierl, 1996) y por ello resulta ser muy útil en la evaluación de secuencias de enseñanza, valoración del currículo y análisis de documentos curriculares (Tatsuoka, 2009).

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

total en la prueba y en q por los niveles de dominio de atributos. Si los valores de R^2 son significativos, entonces se infiere que la lista de atributos es completa y que la matriz de incidencia es correcta (Dogán, 2006). La matriz Q es un modelo cognitivo que preserva las relaciones ya descritas entre atributos en la matriz R y en la jerarquía de atributos presentada. Sin embargo, la matriz Q agrega información sobre cuáles ítems o secuencias de ítems permiten medir patrones de atributos particulares.

Dado que la matriz Q es una matriz de atributos por ítems, su propósito es facilitar la construcción de un conjunto de ítems que sea relevante para el diagnóstico de concepciones que resulten de falta de conocimiento o de la incomprensión de uno o varios atributos. La matriz de incidencia tiene un significado crítico para el desarrollo de pruebas, pues representa las especificaciones cognitivas para su elaboración. Lo que el MRER propone es categorizar a los estudiantes según las secuencias de dominio de atributos, a partir de las respuestas que dan a los ítems de una prueba en particular; de esta manera, las descripciones que brinda del desempeño de los examinados son más enriquecedoras, sobre todo en relación con el mejoramiento de la enseñanza y el aprendizaje (Katz, Martínez, Sheehan & Tatsuoaka, 1993).

Clasificación de examinados según estados de conocimiento

La siguiente etapa del MRER es traducir la información que brinda la matriz Q , en una forma que pueda ser comparada con *patrones individuales observados de respuesta*, los cuales son vectores de unos y ceros que representan cómo un examinado se desempeñó en una prueba. Por ejemplo, el patrón (0 1 0 1 1) indica que el estudiante respondió correctamente los ítems 2, 4, 5 y 6, pero que respondió incorrectamente los ítems 1 y 3.

Para generar los posibles *patrones ideales de respuesta*, que podrían obtenerse de la aplicación de una variedad de reglas, se utiliza un software creado para tal propósito. La aplicación consistente de todos los atributos permitirá obtener las respuestas correctas para toda la prueba. Por otra parte, la apli-

cación consistente de solo algunos atributos dará como resultado un patrón específico de respuesta compuesto por unos y ceros. De esta manera, si un estudiante aplica una combinación específica de atributos para todos los ítems de una prueba, su patrón de respuesta corresponderá exactamente con uno de los patrones ideales generados por el software (Tatsuoka, 1995).

Existen inconsistencias que hacen que los patrones observados de respuesta se desvíen de los patrones ideales de respuesta. Una es que cuando un examinado aplica una regla incorrecta (i.e., no domina un atributo), no necesariamente la aplica en toda una prueba (Birenbaum & Tatsuoaka, 1983). Otra causa de este desajuste se debe a errores por descuido, incertidumbre y fatiga, entre otros, los cuales son considerados errores aleatorios.

Los patrones de respuesta generados a partir de estas inconsistencias en la aplicación de una regla se denominan *patrones difusos de respuesta* y se agrupan alrededor de los patrones ideales formando conglomerados. En el contexto del MRER, se puede llevar a cabo una variedad de análisis de patrones de respuesta, que ofrecen información sobre el significado cognitivo de los patrones difusos de respuesta, pues no se consideran simplemente como patrones erróneos, sino que pueden corresponder a modelos de respuesta alternativos a los ítems y que, en literatura educativa, se conocen como procedimientos erróneos (*bugs*) o concepciones equivocadas (*misconceptions*).

Sin embargo, la variedad de respuestas observadas que forman estos conglomerados (constituidos por patrones ideales y patrones difusos), complica la clasificación de examinados, pues no se pueden asociar exactamente las respuestas observadas con los patrones ideales de respuesta. Para ello, el MRER clasifica los patrones observados en *estados de conocimiento*, los cuales son patrones binarios de atributos que expresan el dominio de tales atributos. La cantidad de posibles estados de conocimiento está dada por 2^k , donde k es el número de atributos. El MRER utiliza una función booleana descriptiva, para relacionar algebraicamente los estados de

conocimiento con los patrones ideales de respuesta (Tatsuoka, 1991).

Considérese la matriz,

$$Q_1 = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

esto es, hay tres atributos y cinco ítems (Im, 2007). El número de *posibles* estados de conocimiento es $2^3 = 8$. Por medio de una función booleanas, se generan los posibles estados de conocimiento y sus correspondientes patrones ideales de respuesta según la matriz Q_1 . Dos o más estados de conocimiento pueden tener asociados el mismo patrón ideal de respuestas (Im, 2007) por las posibles relaciones de inclusión entre atributos. Una vez que se han identificado los patrones ideales de respuesta y que se han generado los estados de conocimiento implicados por una matriz Q , el siguiente paso en el MRER es asociar las respuestas observadas de los

examinados con los patrones ideales, con el fin de clasificarlos en un estado de conocimiento determinado y poder hacer valoraciones diagnósticas a nivel cognitivo para los estudiantes.

Para lograr lo anterior, se representa el conjunto de patrones ideales en un espacio cartesiano conocido como *espacio de representación de reglas* (ERR) (Tatsuoka, 1983; Tatsuoka & Tatsuoka, 1985), cuyos puntos son de la forma (θ, ζ) (Tatsuoka & Linn, 1983; Birenbaum, Kelly & Tatsuoka, 1993; Gierl, 1996; Gierl, Leighton & Hunka, 2000; Rupp, 2007). De esta manera, los puntos que se grafican y que corresponden a los patrones ideales tienen la forma (θ_t, ζ_t) , con $t = 1, \dots, T$ (T es el número de patrones ideales de respuesta) y son llamados *centroides*. Como se ve, en la representación se utiliza el nivel de aptitud q de la teoría de respuesta al ítem en combinación con un nuevo parámetro, denotado por z , llamado *índice extendido de precaución* (Tatsuoka, 1985). Este índice mide la atipicidad de los patrones de respuesta; esto es, estudiantes de alta aptitud que

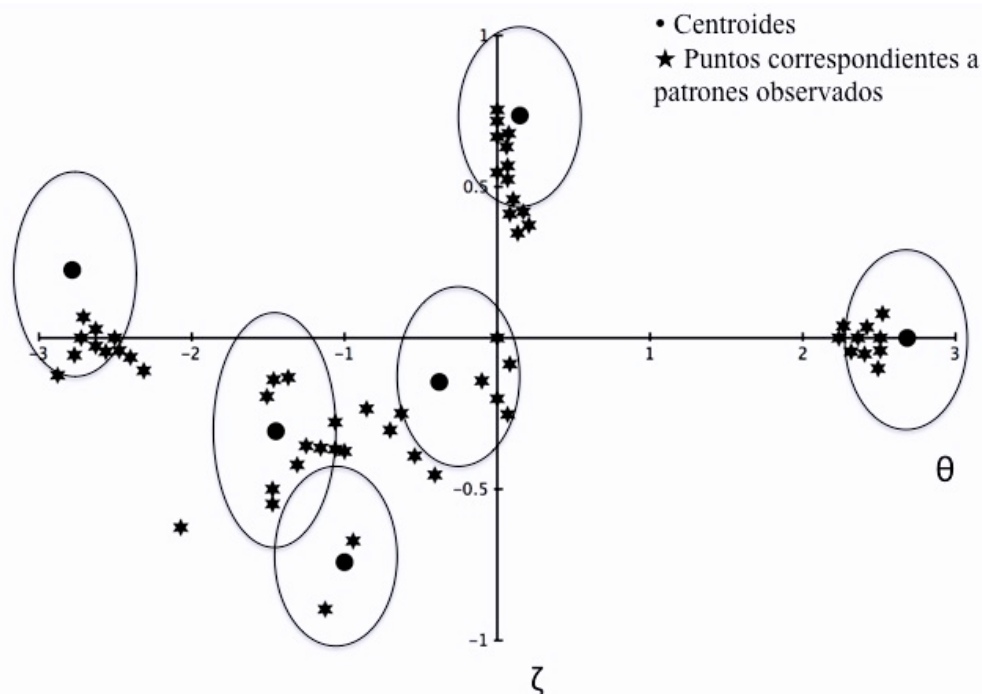


Figura 2. Elipses conformadas hipotéticamente luego de aplicar el punto de corte para la distancia de clasificación en un ERR bidimensional. Fuente: adaptación de Im (2007).

no acierten algunos ítems fáciles, como estudiantes de baja aptitud que acierten algunos ítems difíciles, tendrán valores positivos de z cada vez mayores. En el ERR también se representan las respuestas observadas de los examinados haciéndolas corresponder –por medio de una función booleana– con puntos de la forma (θ_i, ξ_i) , con $i = 1, \dots, N$ (N es el número de examinados). En la Figura 2 se presenta un ejemplo hipotético que muestra la representación gráfica de centroides y puntos correspondientes a patrones observados.

Los puntos correspondientes a patrones observados de respuesta se conglomeran en torno a los centroides y conforman elipses de probabilidad que representan los algoritmos o conceptos erróneos utilizados por los estudiantes. La medición de la proximidad de los patrones de respuesta de los estudiantes con respecto a patrones de respuestas del conglomerado al cual pertenecen, resulta de gran utilidad para diagnosticar la aplicación de reglas erróneas por parte de los examinados. En particular, los puntos ubicados en el extremo izquierdo del espacio tienen valores negativos de z , lo cual quiere decir que los patrones de respuesta a los ítems correspondientes a estos puntos tienen mayor probabilidad de tener puntuaciones de 1 para los ítems fáciles y de 0 para los ítems más difíciles; los puntos ubicados en el extremo opuesto mostrarán una tendencia contraria a la ya descrita (Tatsuoka, 2009).

En la cuarta etapa del MRER, se debe localizar cada estudiante en el espacio multidimensional e interpretar los resultados de la clasificación. Dado que los patrones observados de respuestas pueden diferir de los patrones ideales de respuestas, entonces la ubicación de los examinados se decide utilizando la prueba estadística conocida como distancia de Mahalanobis (D^2). De esta manera, para cada patrón observado de respuestas, se calcula D^2 para todos los patrones ideales cercanos, con lo que cada examinado se puede clasificar en un grupo asociado a un estado cognitivo.

Consideraciones finales

Tendencias recientes de investigación en el área de la evaluación educativa se han enfocado en el desa-

rollo de modelos que especifiquen las estructuras de conocimiento y las habilidades de pensamiento que se requieren para responder los ítems de una prueba. Dichos modelos permiten interpretar las puntuaciones de los examinados de una manera diferente a la que tradicionalmente se ha empleado, pues se recurre a las habilidades cognitivas que subyacen al pensamiento de los estudiantes; además, sirven como guía para los procesos de enseñanza, sobre todo cuando el conocimiento y las habilidades especificados en el modelo se han identificado como aspectos débiles para los estudiantes.

Una de las ventajas del MRER, un modelo que combina principios de la psicología cognitiva con análisis psicométricos tradicionales, consiste en que permite reanalizar datos de pruebas ya existentes y con ello contribuye a la obtención de evidencias de validez de constructo. De acuerdo con Chen (2006), el MRER se puede utilizar para validar el modelo cognitivo subyacente al rendimiento en una prueba, para clasificar examinados en sus estados de conocimiento, y para comparar y contrastar diferencias entre grupos en términos del dominio de atributos cognitivos. A pesar de no ser ni el primero ni el único en su género, el MRER es considerado un caso ejemplar de cómo los datos modelados con métodos psicométricos pueden ser utilizados en la elaboración de inferencias sobre los procesos que utilizan los estudiantes por medio de sus secuencias de respuestas a los ítems de una prueba (Gierl, 1997). De acuerdo con Guerrero (2001), el método desarrollado por Tatsuoka proporciona un poderoso enfoque para el estudio de las estructuras de conocimiento y las estrategias que los estudiantes utilizan para resolver problemas. En consecuencia, el método brinda insumos de gran relevancia a docentes, estudiantes, especialistas en diseño de planes de mejoramiento, personal encargado de elaboración de pruebas, psicólogos cognitivos y especialistas en psicometría.

Otra de las ventajas del MRER es que impulsa tanto a investigadores como a docentes a identificar cuidadosamente las habilidades cognitivas de los estudiantes y poder así establecer un ordenamiento entre ellas, según se requieran en la resolución de ítems, y con ello brindar información valiosa sobre las fortalezas y debilidades de los estudiantes.

lezas y debilidades de los estudiantes en términos de las estrategias cognitivas empleadas en la resolución de problemas (Gierl, 2007; Gierl, Leighton & Hunka, 2000). Además, el MRER también ha mostrado ser de gran utilidad en estudios de funcionamiento diferencial de ítems (Chen, 2009); en modificar y validar un conjunto hipotético de atributos que subyacen al rendimiento en pruebas de comprensión en inglés (Buck & Tatsuoka, 1998); en analizar la equivalencia de constructo de versiones de una misma prueba en dos idiomas distintos (Guerrero, 2001); así como en analizar el rendimiento de estudiantes en pruebas estandarizadas desde una perspectiva cognitiva diagnóstica (Chen, 2006; Dogan, 2006).

Los análisis propios del MRER se pueden enriquecer aún más con el uso de modelos multinivel que permitan analizar la influencia de variables del contexto educativo, tales como estrategias del docente y características del centro educativo (Chen, 2006; Chen, Gorin, Thompson & Tatsuoka, 2008) e incluso en el estudio de análisis de distractores y establecimiento de niveles de desempeño.

Se debe investigar con mayor profundidad el mínimo de examinados que se requiere para llevar a cabo los análisis del MRER, el mínimo de ítems requerido para tales análisis, la proporción ideal entre ítems y atributos y las características de las matrices de incidencia que producen resultados más precisos y confiables (Dogan, 2006). Asimismo, dado que diferentes puntos de corte afectan los análisis del MRER, principalmente las relaciones jerárquicas entre estados de conocimiento, cabe investigar el punto de corte óptimo para establecer el dominio de atributos (Chen et al., 2008). Finalmente, es relevante estudiar la eficiencia del MRER en conjunto con enfoques basados en el funcionamiento diferencial de los ítems para validar los atributos que componen la matriz de incidencia (Chen, 2009). La respuesta a estos interrogantes, sin duda, contribuirá a que, a través del MRER o de modelos semejantes, sea posible inferir con mayor seguridad competencias a partir de pruebas de gran escala que miden conocimiento.

Referencias

- Birenbaum, M., Kelly, A. & Tatsuoka, K. (1993). *Toward a stable diagnostic representation of students' error in algebra* (Technical Report RR-92-58-ONR). Princeton, NJ: Educational Testing Service.
- Birenbaum, M., Kelly, A., Tatsuoka, K. & Gutvirtz, Y. (1994). Attribute-mastery patterns from rule space as the basis for student models in algebra. *Int. J. Hum.-Comput. Stud.*, 40(3), 497-508.
- Birenbaum, M. & Tatsuoka, K. (1983). The effect of a scoring system based on the algorithm underlying the students' response patterns on the dimensionality of achievement test data of the problem solving type. *Journal of Educational Measurement*, 20(1), 17-26.
- Brown, J. S. & VanLehn, K. (1982). Towards a generative theory of "bugs". En T. P. Carpenter, J. M. Moser & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 117-135). Hillsdale, NJ: Erlbaum.
- Buck, G. & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, 15(2), 119-157.
- Buck, G., VanEssen, T., Tatsuoka, K., Kostin, I., Lutz, D. & Phelps, M. (1998). *Development, selection and validation of a set of cognitive and linguistic attributes for the SAT I Verbal: Sentence completion section* (Research Report RR-98-23). Princeton, NJ: Educational Testing Service.
- Chen, Y. (2006). *Cognitively diagnostic examination of Taiwanese mathematics achievement on TIMSS-1999*. Disertación doctoral no publicada, Arizona State University, Arizona, EE. UU.
- Chen, Y. (2009, abril). *A unified approach that combines differential item functioning with the rule-space method for validating cognitive attributes for the TIMSS mathematics items*. Documento presentado en la reunión del National Council on Measurement in Education, San Diego, CA, EE. UU.
- Chen, Y., Gorin, J., Thompson, M. & Tatsuoka, K. (2006, abril). *Verification of cognitive attributes required to solve the TIMSS-1999 mathematics items for Taiwanese students*. Documento presentado en la reunión de la American Educational Research Association, San Francisco, CA, EE. UU.
- 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 en marzo 18 de 2012, de http://www.ierinstitute.org/IERI_Monograph_Volume_01.pdf
- Crandall, B., Klein, G. & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Cambridge, MA: The MIT Press.
- Dogan, E. (2006). *Establishing construct validity of the mathematics subtest of the university entrance examination in Turkey: A rule space application*. Disertación doctoral no publicada, Teachers College, Columbia University, New York, EE. UU.
- Embretson, S. (1990). Diagnostic testing by measuring learning processes: Psychometric considerations for dynamic testing. En N. Frederiksen, R. Glaser, A. Lesgold & M. Shafto (Eds.), *Diagnostic monitoring of skills and knowledge acquisition* (pp. 407-432). Mahwah, NJ: Erlbaum.
- Gierl, M. (1996). *An investigation of the cognitive foundation underlying the rule-space model*. Disertación doctoral no publicada, University of Illinois at Urbana-Champaign, Illinois, EE. UU.
- Gierl, M. (1997). Comparing cognitive representations of test developers and students on a mathematics test with Bloom's Taxonomy. *The Journal of Educational Research*, 91(1), 26-32.
- 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. & Hunka, S. (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, M., Zhou, J. & Alves, C. (2008). Developing a taxonomy of item model types to promote assessment engineering. *Journal of Technology, Learning, and Assessment*, 7(2), 1-50.
- Guerrero, A. (2001). *Cognitively diagnostic perspectives on English and Spanish versions of a test of mathematics aptitude*. Disertación doctoral no publicada, Teachers College, Columbia University, New York, EE. UU.
- Holyoak, K. J. (1995). Problem solving. En D. Osherson & E. E. Smith (Eds.), *Invitation to cognitive science. Thinking* (2nd. ed., Vol. 3, pp. 267-296). Cambridge, MA: The MIT Press.
- Im, S. (2007). *Statistical consequences of attribute misspecification in the rule space model*. Disertación doctoral no publicada, Teachers College, Columbia University, New York, EE. UU.
- Katz, I., Martinez, M., Sheehan, K. & Tatsuoka, K. (1993). *Extending the rule space model to a semantically-rich domain: Diagnostic assessment in architecture* (Technical Report RR-93-42-ON). Princeton, NJ: Educational Testing Service.
- Leighton, J., Gierl, M. & Hunka, S. (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 septiembre de 2007, de <http://www.education.ualberta.ca/educ/psych/crame>
- Lohman, D. (2000). Complex information processing and intelligence. En R. J. Sternberg (Ed), *Handbook of intelligence* (pp. 285-340). New York: Cambridge University Press.
- Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Ohlsson, S. (2007). Psychology is about processes. *Integrative Psychological and Behavioral Science*, 41, 28-34.
- Rupp, A. (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.
- Tatsuoka, K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4), 345-354.
- Tatsuoka, K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics*, 10(1), 55-73.
- Tatsuoka, K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. En N. Frederiksen, R. Glaser, A. Lesgold & M. Shafto

- (Eds.), *Diagnostic monitoring of skills and knowledge acquisition* (pp. 453-488). Hillsdale, NJ: Erlbaum.
- Tatsuoka, K. (1991). *Boolean algebra applied to determination of universal set of misconception states* (Research Report RR-91-44). Princeton, NJ: Educational Testing Service.
- Tatsuoka, K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. En P. Nichols, S. Chipman & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327-359). Hillsdale, NJ: Erlbaum.
- Tatsuoka, K. (2009). *Cognitive assessment: An introduction to the rule space method*. New York: Routledge Taylor & Francis Group.
- Tatsuoka, K. & Linn, R. (1983). Indices for detecting unusual patterns: Links between two general approaches and potential applications. *Applied Psychological Measurement*, 7(1), 81-96.
- Tatsuoka, K. & Tatsuoka, M. (1985). *Bug distribution and pattern classification* (Research Report 85-3-ONR). Urbana, IL: University of Illinois at Urbana-Champaign, Computer-based Education Research Laboratory.
- VanLehn, K. (1989). Problem solving and cognitive skill acquisition. En M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 526-579). Cambridge, MA: The MIT Press.
- Yan, D., Almond, R. & Mislevy, R. (2004). *A comparison of two models for cognitive diagnosis* (Research Report RR-04-02). Princeton, NJ: Educational Testing Service.