

Interpretación y análisis de pruebas educativas y psicológicas con el método *rule space*

Interpretation and Analysis of Educational and Psychological Tests with the Rule Space Method

Álvaro Artavia-Medrano¹

Universidad de Costa Rica, Costa Rica

Resumen. La integración de elementos cognitivos con enfoques psicométricos constituye un campo promisorio en el diseño de pruebas educativas y psicológicas, así como en la interpretación de los resultados que con ellas se obtienen. De igual modo, permiten disponer de evidencias de validez de constructo, puesto que los atributos que se proponen se contrastan con el marco teórico del dominio de conocimiento de interés. En este artículo se destacan los aportes del método *rule space* en las evaluaciones de diagnóstico cognitivo.

Palabras clave. Métodos psicométrico-cognitivos, pruebas educativas, test psicológicos, método *rule space*.

Abstract. The integration of cognitive elements with psychometric approaches is a promising field in the design of educational and psychological tests as well as in the interpretation of their results. Likewise, they provide evidences of construct validity since the proposed attributes are compared with the theoretical framework of the knowledge domain of interest. This article highlights the contributions of rule space method in cognitive diagnostic assessments.

Keywords. Psychometric-cognitive methods, educational tests, psychological tests, rule space method.

¹Álvaro Artavia-Medrano. Escuela de Formación Docente, Facultad de Educación, Universidad de Costa Rica. Dirección Postal: 11501-2060, San José, Costa Rica. Email: alvartavia@gmail.com



Introducción

Los enfoques tradicionales para la interpretación de resultados de pruebas educativas y psicológicas solo ofrecen estimaciones externas del desempeño de las personas en las pruebas (puntuación total o estimaciones de aptitud, por ejemplo), en vez de información interna acerca del conocimiento y las habilidades que poseen las personas como evidencia de sus respuestas a los ítems de una prueba.

Los modelos psicométricos tradicionales, como la Teoría Clásica de los Test (TCT), asumen que el error de medición se distribuye normalmente y de manera igual para todos los niveles de puntuación, lo que impide reconocer diferencias entre personas con distintos patrones de respuesta (Embretson, 1999). Las principales limitaciones de la TCT radican fundamentalmente en el hecho de que no brinda estimaciones invariantes para los parámetros de los ítems y la aptitud de las personas. Hambleton, Swaminathan y Rogers (1991) mencionan la poca utilidad de la TCT en la elaboración de pruebas para poblaciones diferentes y el diseño de bancos de ítems, específicamente por problemas eventuales de comparabilidad, así como el hecho de estar más orientada a la prueba como un todo en vez de a los ítems que la constituyen, por lo que no es posible predecir cómo rendirá una persona o un grupo de personas en un ítem dado, ni diseñar pruebas con ciertas características para determinadas poblaciones.

Posteriormente se desarrollaron propuestas falsables, esto es, que permiten someter a prueba el hecho de que un modelo determinado resulte apropiado o no para un conjunto particular de datos, es decir, evaluar si el modelo predice o explica adecuadamente los datos. Una de esas propuestas es la Teoría de Respuesta a los Ítems (TRI) que plantea que las estimaciones de las puntuaciones de las personas en los rasgos latentes se utilizan para explicar la puntuación que tendrá cada persona en un ítem o en una prueba completa, por lo que la variable independiente es el atributo o rasgo y la variable dependiente es la respuesta al ítem.

Por su parte, Mislevy (2006) afirma que

La TRI no es una teoría acerca de cómo las personas responden ítems, sino que abre la puerta para elaborar

modelos que lo hacen. En la Teoría Clásica de los Test, las afirmaciones y los datos en los argumentos sustantivos son el punto de partida para las variables en el modelo probabilístico (...) La diferencia es que en la TRI, se modela el nivel de tareas individuales. Aunque la información en varias tareas se sintetiza en afirmaciones acerca de una competencia simple y total, los modelos de granularidad más fina brindan una consideración más cuidadosa del significado de las puntuaciones y la detección de patrones de respuesta que entran en conflicto con la presunción de una sola competencia (p. 266).

Actualmente la mayoría de las pruebas psicológicas y educativas se continúan basando en modelos psicométricos tradicionales, los cuales tienen una limitada conexión con los procesos, las estrategias y las estructuras de conocimiento involucradas en la resolución de ítems.

Este artículo tiene como propósito fundamental resaltar los aportes de la integración entre la psicometría y la psicología cognitiva, específicamente, con un ejemplo basado en la utilización del método rule space (Tatsuoka, 1990; 2009), desarrollado por Artavia-Medrano (2014). Para ello, el documento se ha organizado en cuatro secciones, en la primera de ellas se justifica la necesidad de vincular los principios de la psicología cognitiva con la psicometría para una mejor comprensión de la información que brindan las pruebas. En la segunda se explican los aspectos más relevantes del método rule space, uno de los modelos desarrollados para la evaluación de diagnóstico cognitivo. En la tercera, se detalla un ejemplo concreto de la utilidad del rule space en pruebas de gran escala. Finalmente, se exponen algunas consideraciones finales y posibilidades futuras de investigación.

Psicometría y psicología cognitiva: una vinculación necesaria

De acuerdo con Messick (1995), los ítems de una prueba no solo deben cumplir con criterios tradicionales tales como una dificultad apropiada o altos niveles de discriminación, sino también estar justificados en aspectos relevantes del constructo y de los procesos cognitivos involucrados en su resolución.

Diversas investigaciones realizadas por Tatsuoka (1983), Messick (1984), Bejar (1984), Nichols

(1994), Pellegrino, Baxter y Glaser (1999), Leighton, Gierl y Hunka (2002), Mislevy, Steinberg y Almond (2003), Gorin (2006), Rupp, Templin y Henson (2010), entre otros, han planteado la necesidad de combinar principios de la psicología cognitiva con métodos psicométricos para identificar la forma en que las personas organizan su conocimiento, diagnosticar las concepciones erróneas que evidencien y con ello proponer acciones para el mejoramiento de los aprendizajes.

Para Snow y Lohman (1989) “La evidencia de la psicología cognitiva sugiere que el desempeño en una prueba está constituido por complejos conjuntos de acciones de procesamiento de la información que se adaptan a los requerimientos de la tarea durante el desempeño” (p. 317). Por ello, tales autores identifican cuatro posibilidades de los beneficios de integrar los avances de investigaciones en el ámbito cognitivo con los modelos psicométricos empleados para la interpretación de resultados de pruebas: a) brindar una nueva forma de pensar y comprender las puntuaciones, b) describir los constructos medidos en una prueba mediante componentes cognitivos tales como representación, conocimiento, selección de estrategias y procesos de resolución, c) explicitar los supuestos psicológicos utilizados para diseñar pruebas y asignar puntuaciones, y d) permitir la medición de aptitud, aprendizaje, enseñanza y rendimiento con teorías más unificadas y con mayor poder de predicción en la medición de componentes cognitivos.

De esta manera, la vinculación entre la psicología cognitiva y la psicometría encuentra su principal manifestación en el desarrollo de modelos de diagnóstico cognitivo, cuyo propósito fundamental es identificar fortalezas y debilidades para hacer inferencias acerca de las habilidades de las personas en la resolución de diversas situaciones. Según Gierl, Leighton y Hunka (2007) “una evaluación de diagnóstico cognitivo requiere de un enfoque cognitivo de procesamiento de la información para modelar la psicología del rendimiento en una prueba, pues las inferencias sobre las puntuaciones se orientan específicamente hacia las habilidades cognitivas de las personas” (p. 242).

El análisis del procesamiento de la información propuesto por Newell y Simon (1972) proporciona el tipo de estudio detallado de las capacidades humanas que permite la continuidad desde los modelos estables de diferencias individuales que proporcionan las teorías psicométricas, hasta la identificación de los procesos, los contenidos y las representaciones individuales empleadas por las personas en la resolución de tareas específicas.

La perspectiva teórica del procesamiento de la información ha recibido diversas críticas, sobre todo en lo que se refiere a adoptar los programas de computación como metáfora del funcionamiento cognitivo humano, esto es, aceptar la analogía entre la mente humana y el funcionamiento de una computadora (Pozo, 2002). Asimismo, la psicología cognitiva actual ha dado otras explicaciones al tratamiento de la información, sobre todo desde un punto de vista sociocultural, conexionista y neurocientífico. No obstante, el procesamiento de la información sigue siendo adecuado para el estudio de la naturaleza, el contenido y el empleo de mucha de la información que las personas utilizan (Rupp & Mislevy, 2007), pues posibilita analizar detalladamente las estructuras y los procesos mentales asociados a la resolución de los ítems de una prueba en particular. Por ello, la perspectiva psicométrica y la cognitiva se pueden ver como complementarias entre sí y desarrollarse mejor conjuntamente en vez de considerarlas como entidades separadas (Sternberg, 1986).

Consecuentemente, este enfoque permite la valoración de una serie de aspectos cruciales para el aprendizaje escasamente considerados en las pruebas tradicionales, como los conocimientos o las ideas previas de las personas en relación con dominios específicos, así como las estrategias de aprendizaje de que disponen o sus capacidades metacognitivas. Con ello, se puede mejorar sustancialmente la utilidad diagnóstica de una prueba.

En particular, Nichols (1994) utilizó el término “evaluación cognitiva diagnóstica”, para resaltar la integración de la psicología cognitiva en el diseño y la interpretación de los resultados de pruebas, así como en la elaboración de diagnósticos sobre

las características de las personas en cuanto a sus procesos cognitivos.

Con el propósito de diseñar pruebas a partir de un enfoque de diagnóstico cognitivo, Nichols (1994) propuso cinco pasos: a) el desarrollo de un modelo o teoría que permita identificar las variables de los ítems que se refieren a estructuras de conocimiento y procesos cognitivos particulares, b) la selección de un diseño de observación y medición, en el que los ítems utilizados permitan que las personas respondan utilizando los procesos cognitivos y las estructuras de conocimiento ya identificadas, c) las decisiones acerca del entorno y el contexto en que se desarrollará la prueba y su incidencia en el desempeño de las personas, d) la calificación de las respuestas, y e) la revisión del diseño para acumular evidencia que fundamente la teoría seleccionada o bien, sugiera cambios a partir de los resultados de la evaluación.

De esta manera, al tener presente las orientaciones dadas para su correcta elaboración, se reconoce que la evaluación de diagnóstico cognitivo brinda información sobre las estrategias utilizadas por las personas en la resolución de situaciones, las relaciones entre conceptos que ellas perciben y los principios propios de un dominio evaluado. La especificidad de la información está dada en términos explicativos de por qué las personas respondieron de una manera determinada, es decir, cómo vincular el desempeño de una de ellas en una prueba con las inferencias que se puedan hacer de sus fortalezas y debilidades en el ámbito cognitivo (Leighton & Gierl, 2007; 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 evidenciadas en la resolución de problemas con las interpretaciones que se puedan hacer sobre su desempeño.

En el campo de la medición, 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, Brito & Gotzmann, 2009; Leighton & Gierl, 2007).

Los modelos cognitivos son indispensables en la evaluación de diagnóstico cognitivo 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.

En la literatura está suficientemente documentado que muchas de las aplicaciones de modelos complejos de diagnóstico se llevan a cabo mediante un análisis post hoc, también llamado ajuste posterior o *retrofitting*, lo que ha recibido críticas por las limitaciones que puedan tener en brindar una adecuada clasificación para las personas (Gierl & Cui, 2008). No obstante, Roussos, DiBello, Henson, Jang y Templin (2010) afirman que este tipo de análisis se realizan

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

Uno de los primeros esfuerzos en representar las habilidades cognitivas requeridas para resolver ítems en pruebas educativas y psicológicas e integrarlas con enfoques psicométricos, lo constituyó el modelo rule space de Tatsuoka (1983, 1990, 2009), que se describirá en la siguiente sección.

El método rule space: descripción de atributos y clasificación para el diagnóstico

Con el propósito de enriquecer las interpretaciones de los modelos básicos de TRI, se propuso la incorporación de información externa a los parámetros de dichos modelos, que tuvieran origen en procesos de respuesta fundamentados en la psicología cognitiva.

Una de esas propuestas se desarrolló en un programa de investigación dirigido por Kikumi Tatsuoka (1983, 1990, 2009) y que originó la “metodología rule space” en

la que se combina el modelado de variables latentes propio de la TRI con el análisis bayesiano de conglomerados.

El trabajo en dicha metodología estuvo motivado en sus orígenes por el análisis de errores en áreas o contenidos sumamente específicos de la Matemática, tales como álgebra, números negativos, fracciones y exponentes (Birenbaum & Tatsuoka, 1993). El propósito de la utilización del método en tales estudios fue principalmente diagnosticar el funcionamiento de las habilidades de las personas en la resolución de problemas y detectar patrones inusuales de respuesta al responder ítems propios de aritmética (Tatsuoka, 1983).

Distintos estudios mostraron la inestabilidad de los errores sistemáticos; además, tales errores resultan ser soluciones tentativas para resolver problemas cuando las personas no tienen las habilidades apropiadas. Incluso, se tienen evidencias de que a menudo las personas no recuerdan cuáles reglas erróneas utilizaron ni las pueden describir unos cuantos segundos después de su uso (Shaw, 1986).

Debido a que la consideración de componentes cognitivos relevantes resulta de mayor estabilidad y relevancia para procesos educativos y psicológicos, el método rule space se propuso como un enfoque probabilístico que se basa en un análisis de los requerimientos cognitivos en una tarea y que se denominan atributos. De acuerdo con Birenbaum y Tatsuoka (1993) “un atributo de una tarea es una expresión de la dimensión subyacente de la tarea que se requiere para completarla exitosamente. Los atributos pueden incluir procedimientos, heurísticas, estrategias, habilidades y otros componentes del conocimiento” (p. 256). Adicionalmente, Birenbaum, Kelly y Tatsuoka (1992) afirman que los atributos pueden incluir la adopción de una estrategia particular.

El método rule space se desarrolla en cuatro etapas: a) identificación de las habilidades cognitivas requeridas para resolver los ítems de una prueba, b) generación de los patrones ideales de respuestas, c) proyección de los patrones de respuesta de las personas y de los patrones ideales en un espacio bidimensional, y d) clasificación de estudiantes en grupos de dominio de atributos.

La primera etapa del rule space la constituye la identificación de los atributos de los ítems, la cual es un proceso interactivo que inicia con la elaboración de una lista preliminar a partir de la investigación de literatura en el dominio específico de conocimiento, el análisis de protocolos verbales y escritos de estudiantes, así como de la participación de docentes o especialistas con experiencia, entre otros.

Con los atributos identificados por ítem, se genera la matriz de incidencia o matriz Q (Tatsuoka, 1983, 1990, 2009), la cual está compuesta por unos y ceros (los cuales indican la presencia o no del atributo en el ítem, respectivamente). La matriz Q tiene k filas y j columnas, donde k es la cantidad de atributos y j es la cantidad de ítems. Para la validación de la matriz de incidencia, se llevan a cabo análisis de regresión múltiple en los que la variable dependiente es la dificultad de los ítems (estimada con algún modelo de TRI) y las variables independientes son los atributos; se considera aceptable un coeficiente de determinación mayor o igual que 0.80.

La segunda etapa del método es la generación de “estados de conocimiento”, los cuales son vectores de atributos que van desde no dominar ninguno de los atributos de la prueba (en cuyo caso se tendrá un vector compuesto únicamente por j ceros), hasta el vector que represente que se dominan todos los atributos correspondientes a los ítems de la prueba (un vector compuesto únicamente por j unos). La cantidad posible de estados de conocimiento está dada por 2^k , donde k es la cantidad de atributos que componen la matriz Q .

El método rule space, emplea una función booleana descriptiva (Tatsuoka, 2009) para asociar los estados de conocimiento (que son posibles combinaciones de atributos según las especificaciones de la matriz Q) con los patrones de respuestas de las personas. Para ello se parte del supuesto básico de que una persona acierta un ítem si y solo si domina todos los atributos involucrados en la resolución de dicho ítem. Por esta razón se dice que el rule space es un método no compensatorio (Rupp, Templin & Henson, 2010).

En la tercera etapa del método, se hace una representación gráfica de los patrones ideales de respuestas a los ítems de una prueba en términos de dos variables: θ (theta) y ζ (zeta). La primera de ellas corresponde al nivel de aptitud o habilidad estimada según algún modelo de TRI. Se espera que una persona de alta habilidad tenga un patrón de respuestas con muchos unos y pocos ceros y, por el contrario, una persona en el extremo inferior del continuo de habilidades tendría un patrón de respuestas con la mayoría de elementos iguales a cero. No obstante, puede suceder que tanto una persona de habilidad alta no acierte algunos ítems fáciles, como una de habilidad baja que acierte algunos ítems difíciles, por lo que sus respuestas se considerarán inusuales o atípicas. Para ello, Tatsuoka (1987) introdujo el índice extendido de precaución, denotado por ζ y que constituye la segunda dimensión que el método rule space utiliza. Este índice puede asumir valores negativos, lo cual indica que los patrones de respuesta a los ítems correspondientes tienen mayor probabilidad de tener puntuaciones de 1 para los ítems fáciles y de 0 para los ítems más difíciles; valores positivos para ζ mostrarán una tendencia contraria a la ya descrita, es decir, por lo general unos para los ítems más difíciles y ceros para los más fáciles (Tatsuoka, 2009).

Finalmente, en la cuarta etapa, para cada patrón observado de respuestas, se calcula el valor de D^2 (distancia de Mahalanobis) entre dicho patrón y todos los patrones esperados según las especificaciones de la matriz Q , para así clasificar a cada persona en un grupo determinado, conocido como estado de conocimiento.

Como θ y ζ siguen una distribución normal bivariada (Tatsuoka, 1985), la distancia de Mahalanobis sigue una distribución ji cuadrada (χ^2) con dos grados de libertad en un espacio bidimensional. Si el cálculo del valor de D^2 es menor que el valor crítico de ji cuadrada, entonces el patrón de dominio de atributos (estado de conocimiento) asociado al patrón esperado de respuesta se acepta para clasificar a las personas.

De acuerdo con Im (2007), al utilizar un valor de ζ es posible obtener un punto de corte para las distancias

de clasificación. Así, si al calcular D^2 entre la ubicación de una persona y un centro se obtiene un resultado menor que el punto de corte para la distancia, entonces el estado de conocimiento correspondiente al centro se puede aceptar para dicha persona. De hecho, el punto de corte para la distancia de clasificación correspondiente al 90% de la distribución ji cuadrada con dos grados de libertad es 4.5. Esto quiere decir que solo aquellas personas cuyas distancias a cualquier centro sean menores que el punto de corte pueden ser clasificadas según el rule space.

El dominio de atributos para cada persona se expresa en forma probabilística mediante una combinación de los estados de conocimiento generados según las especificaciones de la matriz Q ponderados con las probabilidades posteriores calculadas con las reglas bayesianas de decisión; a esto se le conoce como probabilidad de dominio de atributos. Una vez que estos vectores de probabilidad se calculan, se pueden inferir características de diagnóstico cognitivo tanto a nivel grupal como individual y con ello finaliza el procedimiento.

Cuando se tienen k atributos, la cantidad de posibles estados previos de conocimiento es 2^k . No obstante, las relaciones entre los atributos al estar involucrados en diversos ítems hacen que no todos esos estados previos estén acordes con las especificaciones de la prueba según la matriz de incidencia, por lo que es posible reducir su cantidad, debido a dichas condiciones iniciales. Los estados de conocimiento similares se agrupan tomando en consideración las probabilidades de dominio de atributos obtenidas con el rule space. Para ello, se llevan a cabo análisis de conglomerados de K medias, el cual parte del análisis de casos individuales para ir agrupando casos hasta llegar a la formación de conglomerados homogéneos.

Al tomar como base los patrones de dominio de atributos, es posible establecer una red de relaciones jerárquicas con un conjunto de estados agrupados de conocimiento. Dicha jerarquía permite describir el orden en que las personas adquieren las habilidades cognitivas identificadas y se les conoce como “trayectorias de aprendizaje” (Tatsuoka, 2009).

Para elaborar una red entre estados de conocimiento se utilizan principios de relaciones de inclusión y de teoría de grafos. Con estas bases matemáticas, es posible establecer que un par de estados agrupados de conocimiento tienen una relación de orden si cada una de las componentes del vector binario de dominio es mayor o igual que su respectiva componente en el otro vector.

En la siguiente sección, a modo de ejemplo, se describen parte de los hallazgos de la investigación de Artavia-Medrano (2014), quien se basó en una prueba de rendimiento académico en Matemática compuesta por 55 ítems, aplicada en estudiantes costarricenses de undécimo año a nivel nacional. Esta prueba es de altas consecuencias para el estudiantado, pues constituye uno de los requisitos para concluir la educación secundaria en Costa Rica.

Inferencias para el diagnóstico cognitivo: un ejemplo

Artavia-Medrano (2014) desarrolló un modelo de diagnóstico cognitivo en Matemática. Para ello, empleó una prueba de gran escala compuesta únicamente por ítems de selección única, cada uno con cuatro opciones de respuesta. En cuanto a los temas medidos en la prueba y el peso porcentual con que aparecen, se tiene: álgebra (20%), funciones (27%), función exponencial y función logarítmica (18%), geometría (20%) y trigonometría (15%). Para su estudio, contó con un conjunto de datos codificados de manera dicotómica (1 en caso de acierto, 0 en caso contrario).

El coeficiente alfa de Cronbach reportado es de 0.86, como evidencia de una alta consistencia interna. Para la estimación de los parámetros de los ítems, se seleccionó el modelo de Rasch, según el cual la respuesta a un ítem depende solo de la aptitud de la persona y de la dificultad del ítem; ambas se estiman en una misma escala y cada una de esas estimaciones tiene un grado de error asociado a ellas, el cual decrece a medida que la dificultad y la aptitud aumentan (Bond & Fox, 2001). Los cálculos con el modelo de Rasch permiten reportar valores de θ que variaron de -2.41 a 3.19, con un promedio de 0 y una desviación estándar de 1.20.

Por las características de la base de datos con la que se contó para desarrollar la investigación, se utilizó muestreo aleatorio simple y así se calculó un tamaño de muestra para estimar una proporción con el 95% de confianza. Para ello, se utilizó el valor 0.5 como desviación estándar máxima en el caso de una proporción, y 0.05 como error de muestreo máximo permisible, con lo que se obtuvo $n = 384$.

En la tabla 1 se describen los atributos correspondientes a los ítems de la prueba en estudio, que fueron codificados y con ello se formuló la matriz Q .

Con el fin de obtener evidencias de validez para la matriz Q , se realizaron diversos análisis de regresión múltiple. En cuanto a los ítems, lo que se pretende es explicar en qué medida la varianza en la dificultad de los ítems se explica por los atributos involucrados en ellos. Los 18 atributos propuestos explican más del 79% de la varianza en la dificultad de los ítems estimada según el modelo de Rasch ($R^2 = 0.861$; $R^2_{\text{adj}} = 0.792$).

De igual modo, se estimó el poder explicativo de las probabilidades de dominio de atributos para predecir el desempeño o rendimiento en la prueba. Para este caso, se obtuvo que el 90% de la varianza en las puntuaciones totales de la prueba se explican por las probabilidades de dominio de atributos ($R^2 = 0.901$; $R^2_{\text{adj}} = 0.896$).

Una vez que se ejecutaron los análisis propios del método rule space, se logró una tasa de clasificación del 98.4% para los patrones observados de respuesta en uno o más de los estados de conocimiento posibles según el método. Esta tasa de clasificación es considerada muy alta. Los patrones de respuestas de las personas no clasificadas es porque resultan completamente inconsistentes con los estados de conocimiento generados y por ello sus respuestas no son comparables ni se aproximan a los patrones esperados de respuestas según la matriz Q propuesta. En el rule space se asume que estas discrepancias se deben principalmente a errores aleatorios o sistemáticos que cometen las personas al resolver los ítems de la prueba (Tatsuoka, 2009).

A pesar de la reducción hecha con los análisis del método rule space, la cantidad de estados de conocimiento puede continuar siendo muy alta debido al número de atributos involucrados en la prueba. Por ello, se hace necesario establecer una cantidad de estados de conocimiento que brinde una información diagnóstica con mayor significado y que sea más interpretable.

Artavia-Medrano (2014) seleccionó la cantidad de 12 conglomerados por considerar que refleja mejor el desempeño o logro de los atributos y porque de ella es posible obtener una relación jerárquica entre los estados conglomerados de conocimiento cuya interpretación tenga mayor sentido educativo. La distancia promedio entre las personas y el centro del conglomerado es 0.74

con una desviación estándar de 0.27 y su rango varía desde 0.22 hasta 1.85.

Con el propósito de entender la importancia de cada atributo en la separación por los 12 conglomerados, se llevó a cabo un análisis de varianza (ANOVA) de un solo factor. Los resultados, que se presentan en la tabla 2, indican cuáles atributos contribuyen más a la solución escogida de los estados conglomerados de conocimiento. Así, los atributos con valores de F grandes, proporcionan mayor separación entre conglomerados; en este caso, la utilización de propiedades de potencias o logaritmos (AT13) y la utilización de métodos de factorización (AT1).

Tabla 1

Atributos para los ítems de la prueba

Atributo	Descripción
AT1	Utilización de métodos de factorización
AT2	Revisión de opciones
AT3	Simplificación de expresiones
AT4	Resolución de ecuaciones o desigualdades
AT5	Traducción y formulación de expresiones para resolver un problema
AT6	Interpretación de la información contenida en el enunciado
AT7	Determinación o cálculo de conceptos básicos de funciones
AT8	Verificación de proposiciones
AT9	Utilización de la información explícita o implícita en una figura
AT10	Distinción entre variables y parámetros
AT11	Utilización de fórmulas
AT12	Determinación o cálculo de conceptos de funciones exponenciales y logarítmicas
AT13	Utilización de propiedades de potencias o logaritmos
AT14	Determinación de elementos o propiedades de una figura plana
AT15	Determinación de elementos o propiedades de dos o más figuras planas o un cuerpo geométrico
AT16	Cálculo de áreas o volúmenes de figuras o regiones
AT17	Establecimiento de equivalencias entre expresiones trigonométricas
AT18	Determinación o cálculo de conceptos de funciones trigonométricas

Nota. Artavia-Medrano (2014).

Tabla 2

Resultados del ANOVA de un solo factor para evaluar las diferencias entre los doce estados conglomerados de conocimiento por atributo

Atributo	Conglomerado		F
	Media cuadrática (gl= 11)	Media Cuadrática (gl=366)	
AT1	1.092	0.007	154.882
AT2	0.710	0.004	189.895
AT3	0.045	0.012	3.737
AT4	0.754	0.006	136.791
AT5	1.139	0.065	17.462
AT6	1.542	0.042	36.570
AT7	2.145	0.049	44.201
AT8	0.254	0.055	4.600
AT9	1.038	0.032	32.455
AT10	0.544	0.052	10.531
AT11	0.104	0.008	12.664
AT12	0.934	0.050	18.507
AT13	3.511	0.007	537.073
AT14	1.010	0.043	23.411
AT15	1.322	0.051	25.872
AT16	0.477	0.054	8.864
AT17	1.640	0.053	31.052
AT18	0.941	0.053	17.599

Nota. $p < .001$ para todos los atributos. Artavia-Medrano (2014).

Puesto que los conglomerados se han elegido para maximizar las diferencias entre los casos, las pruebas F solo se pueden utilizar con una finalidad descriptiva, esto es, no pueden interpretarse como pruebas de la hipótesis de que los centros de los conglomerados son iguales.

Los resultados presentados en la tabla 2 muestran que los valores de F varían desde 3.737 hasta 537.073, esto denota que cada atributo es de utilidad para la separación en estados conglomerados de conocimiento. Por tanto, la escogencia de doce conglomerados se considera una solución interpretable en términos cognitivos.

Una vez que se obtuvieron los conglomerados, se estimó la probabilidad de dominio o logro para cada atributo en cada conglomerado, mediante el promedio de las probabilidades de dominio de atributos para las personas clasificadas en el conglomerado respectivo. En la tabla 3 se muestra la probabilidad media y la cantidad de personas en cada conglomerado.

A modo de ilustración, el ECC1 (estado conglomerado de conocimiento 1) presenta el valor 0.99 como promedio de probabilidad de dominio de los atributos revisión de opciones (AT2) y resolución de ecuaciones o desigualdades (AT4); no obstante, en este conglomerado, se tiene una probabilidad media de 0.21 para el dominio de la determinación de elementos o propiedades de dos o más figuras planas o un cuerpo geométrico (AT15). Para propósitos de brindar información diagnóstica, las 49 personas que se ubican en el ECC1 pueden ser caracterizadas atendiendo a la naturaleza de los atributos con mayor probabilidad media de dominio (por ejemplo, mayores probabilidades en los ítems que requieren algoritmos para su resolución).

Para efectos de interpretaciones con mayor significado educativo, cada uno de los vectores de probabilidad de dominio de atributos de cada conglomerado se transformó en un vector binario, en el que los unos indican el dominio del atributo correspondiente y los ceros, el caso contrario. Con esto se logra un patrón de dominio de atributos para cada estado conglomerado de conocimiento. Se utilizó 0.75 como punto de corte para estas transformaciones.

Por ejemplo, en la primera columna la tabla 3, se evidencia que hay 49 personas que se ubican en el estado conglomerado de conocimientos 1 (ECC1). Para este conglomerado, las personas no dominan: la traducción y formulación de expresiones para resolver un problema (AT5), la verificación de proposiciones (AT8), la utilización de información implícita o explícita en una figura (AT9), la distinción entre variables y parámetros (AT10), la determinación o el cálculo de conceptos de funciones exponenciales y logarítmicas (AT12), la determinación de elementos o propiedades de una figura plana (AT14), la determinación de

elementos o propiedades de dos o más figuras planas o un cuerpo geométrico (AT15), el cálculo de áreas o volúmenes de figuras o regiones (AT16) y la determinación o el cálculo de conceptos de funciones trigonométricas (AT18).

Como se observa en la tabla 4, cuando se consideran patrones binarios, el ECC 1 contiene ceros en dichos

atributos y unos en los demás, pues al superar el valor de corte se consideran como dominados por las personas que integran el conglomerado respectivo. En esa misma tabla se muestran los patrones de dominio ordenados de menor a mayor, junto a la cantidad de atributos dominados y el número de personas clasificadas en cada conglomerado.

Tabla 3

Centros de cada estado conglomerado de conocimiento

Atributo	Estados conglomerados de conocimiento (ECC)											
	1	2	3	4	5	6	7	8	9	10	11	12
AT1	0.98	0.99	1.00	1.00	0.10	0.99	0.99	0.94	0.99	1.00	0.05	1.00
AT2	0.99	0.99	1.00	1.00	0.29	1.00	1.00	0.97	0.99	1.00	0.23	1.00
AT3	0.98	1.00	1.00	1.00	0.89	0.95	1.00	0.98	1.00	0.75	0.86	0.97
AT4	0.99	1.00	0.25	0.25	0.99	0.97	1.00	0.41	1.00	0.07	1.00	1.00
AT5	0.46	0.69	0.15	0.15	0.64	0.77	0.78	0.43	0.82	0.65	0.79	0.30
AT6	0.75	0.41	0.00	0.00	0.57	0.61	0.82	0.90	0.97	0.82	0.98	0.57
AT7	0.78	0.74	0.95	0.95	0.55	0.62	0.19	0.47	0.95	0.95	0.49	0.71
AT8	0.45	0.49	0.35	0.35	0.35	0.45	0.57	0.69	0.57	0.13	0.32	0.46
AT9	0.52	0.96	1.00	1.00	0.66	0.92	0.80	0.37	0.94	0.78	1.00	0.65
AT10	0.51	0.69	0.75	0.75	0.50	0.68	0.65	0.63	0.78	0.61	0.79	0.37
AT11	0.89	0.97	1.00	1.00	0.89	0.96	0.99	0.92	1.00	0.95	1.00	0.87
AT12	0.49	0.42	0.20	0.20	0.39	0.49	0.66	0.83	0.79	0.52	0.91	0.44
AT13	0.98	0.99	0.00	0.00	0.41	0.09	0.99	0.11	1.00	0.02	1.00	0.97
AT14	0.45	0.86	0.85	0.85	0.43	0.85	0.82	0.58	0.79	0.35	1.00	0.60
AT15	0.21	0.72	0.80	0.80	0.48	0.56	0.55	0.16	0.61	0.51	0.77	0.29
AT16	0.52	0.61	0.90	0.90	0.58	0.39	0.53	0.67	0.77	0.58	0.62	0.46
AT17	0.83	0.92	0.60	0.59	0.50	0.86	0.87	0.33	0.89	0.62	0.78	0.22
AT18	0.68	0.83	0.00	0.00	0.53	0.67	0.87	0.53	0.86	0.85	0.75	0.40
Promedio	0.69	0.79	0.60	0.60	0.54	0.71	0.78	0.61	0.87	0.62	0.74	0.63
Cantidad de casos	49	82	14	2	8	18	59	12	87	4	7	36

Nota. Artavia-Medrano (2014).

Con los conglomerados ya formulados, es posible establecer redes de estados de conocimiento, como se muestra en la figura 1 y elaborar trayectorias de aprendizaje conformadas por cadenas de estados de conocimiento. Las trayectorias de aprendizaje permiten definir la manera en que se puede ir mejorando progresivamente en el dominio de habilidades. Algunas de ellas se muestran en la tabla 5. Las trayectorias de aprendizaje mostradas anteriormente inician en EC5, lo cual implica que los tres atributos asociados con dicho estado conglomerado de conocimiento fueron dominados por casi todas las

personas. Tales atributos son: simplificación de expresiones (AT3), resolución de ecuaciones o desigualdades (AT4) y utilización de fórmulas (AT11).

Asimismo, la cadena EC7→EC11→EC9 indica que los atributos: verificación de proposiciones dadas en el enunciado del ítem (AT8), la determinación de elementos o propiedades de dos o más figuras planas o un cuerpo geométrico (AT15) y el cálculo de áreas o volúmenes de figuras o regiones (AT16), resultaron ser los más difíciles de dominar.

Tabla 4

Patrones de dominio de atributos para los estados conglomerados de conocimiento

Atributo	Estados conglomerados de conocimiento (ECC)											
	5	3	8	12	10	6	1	4	2	7	11	9
AT1	0	1	1	1	1	1	1	1	1	1	0	1
AT2	0	1	1	1	1	1	1	1	1	1	0	1
AT3	1	1	1	1	1	1	1	1	1	1	1	1
AT4	1	1	0	1	0	1	1	0	1	1	1	1
AT5	0	0	0	0	0	1	0	0	0	1	1	1
AT6	0	0	1	0	1	0	1	0	0	1	1	1
AT7	0	1	0	0	1	0	1	1	0	0	0	1
AT8	0	0	0	0	0	0	0	0	0	0	0	0
AT9	0	0	0	0	1	1	0	1	1	1	1	1
AT10	0	0	0	0	0	0	0	1	0	0	1	1
AT11	1	1	1	1	1	1	1	1	1	1	1	1
AT12	0	0	1	0	0	0	0	0	0	0	1	1
AT13	0	0	0	1	0	0	1	0	1	1	1	1
AT14	0	0	0	0	0	1	0	1	1	1	1	1
AT15	0	0	0	0	0	0	0	1	0	0	1	0
AT16	0	0	0	0	0	0	0	1	0	0	0	1
AT17	0	0	0	0	0	1	1	0	1	1	1	1
AT18	0	0	0	0	1	0	0	0	1	1	1	1
Atributos dominados	3	6	6	6	8	9	9	10	10	12	13	16
Total	49	82	14	2	8	18	59	12	87	4	7	36

Nota. Artavia-Medrano (2014).

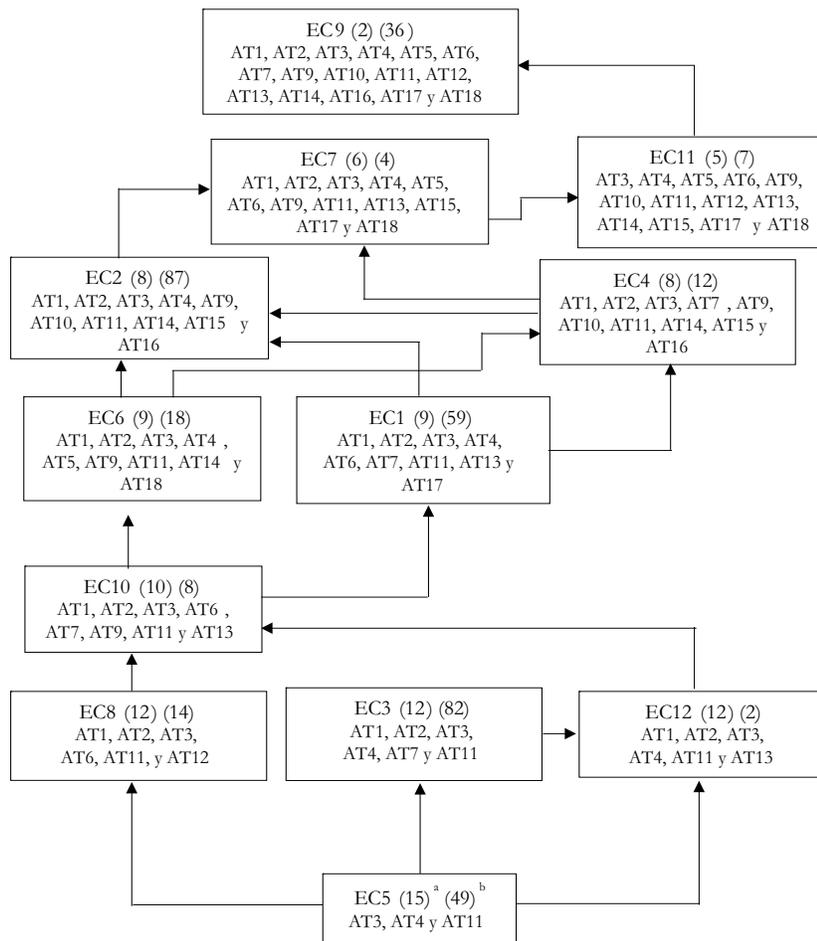


Figura 1. Redes entre estados conglomerados de conocimiento. Elaboración a partir de Artavia-Medrano (2014). ^a El número representa la cantidad de atributos no dominados, por lo que los atributos indicados en cada cuadro son los que se han dominado en cada estado conglomerado de conocimiento. ^b Representa la cantidad de personas clasificadas en ese conglomerado.

Tabla 5

Algunas trayectorias de aprendizaje según estados conglomerados de conocimiento

Trayectoria	Proceso de la trayectoria
1	EC5 → EC3 → EC12 → EC10 → EC6 → EC4 → EC7 → EC11 → EC9
2	EC5 → EC8 → EC10 → EC6 → EC4 → EC7 → EC11 → EC9
3	EC5 → EC12 → EC10 → EC6 → EC4 → EC7 → EC11 → EC9
4	EC5 → EC12 → EC10 → EC1 → EC4 → EC2 → EC7 → EC11 → EC9
5	EC5 → EC3 → EC12 → EC10 → EC1 → EC4 → EC7 → EC11 → EC9
6	EC5 → EC3 → EC12 → EC10 → EC1 → EC4 → EC12 → EC7 → EC11 → EC9

Nota. Artavia-Medrano (2014).

Conclusiones

Tradicionalmente, el desempeño de una persona en una prueba educativa o psicológica se ha definido con base en su puntuación total o estimaciones de su aptitud. Este tipo de apreciaciones permite identificar diferencias entre grupos, pero no dan una explicación del porqué de tales diferencias, ni ofrecen una descripción de su naturaleza u origen.

La identificación del dominio de ciertas habilidades en un campo específico de conocimientos, es posible mediante el establecimiento de atributos cognitivos, considerados como las características subyacentes a los ítems o las habilidades que las personas requieren para un adecuado desempeño en esos ítems.

Dado que los atributos son considerados como componentes de un constructo por medir, la matriz de incidencia refleja la interacción entre los ítems y los atributos identificados, por lo que se le considera un modelo cognitivo para el desempeño de las personas en una prueba dada.

La vinculación entre psicología cognitiva y psicometría se plantea como necesaria en la medida en que permite obtener información diagnóstica con distintos niveles de detalle. Individualmente, es posible saber qué pueden hacer las personas y cuáles estrategias se podrían emplear para mejorar las habilidades que no dominan. En términos grupales, es posible reconocer las diferencias observadas a partir de los patrones de respuesta y del dominio de atributos, lo cual no es factible con los métodos psicométricos tradicionales, en los que personas con la misma puntuación se ubicarían en el mismo punto de la escala.

En particular, el método rule space permite elaborar conglomerados a partir de las probabilidades de dominio de atributos. Estos conglomerados posibilitan establecer redes entre estados de conocimiento y, posteriormente, trayectorias de aprendizaje en las que se evidencia la forma progresiva en que se puede avanzar en el dominio de atributos, dado su ordenamiento jerárquico. De igual modo es posible obtener estas trayectorias para casos individuales, a partir de las relaciones entre estados de conocimiento y su punto

de ubicación en el rule space. Estos conglomerados también permiten corroborar cuáles son los atributos cuyo dominio resultaría más fácil para las personas y las posibilidades que tienen de lograr su dominio, lo cual es de gran significado para diagnósticos cognitivos y planes remediales.

Finalmente, se plantean como sugerencias de investigación, determinar características de la matriz de incidencia para producir resultados más precisos y confiables tanto en el diagnóstico cognitivo como en la clasificación de personas, así como evaluar empíricamente el ajuste entre el modelo propuesto y los datos, con el fin de determinar estrategias de resolución alternativas por parte de las personas o bien, el establecimiento de otros modelos cognitivos.

Referencias

- Artavia-Medrano, A. (2014). *Evaluación cognitiva diagnóstica en Matemática: modelo elaborado con el método rule space para estudiantes costarricenses de undécimo año*. (Tesis doctoral inédita). Universidad de Costa Rica, Costa Rica.
- Bejar, I. (1984). *Educational Diagnostic Assessment*. *Journal of Educational Measurement*, 21(2), 175-189.
- Birenbaum, M., & Tatsuoka, K. (1993). Applying an IRT-Based Cognitive Diagnostic Model to Diagnose Students' Knowledge States in Multiplication and Division with Exponents. *Applied Measurement in Education*, 6(4), 255-268.
- Birenbaum, M., Kelly, A., & Tatsuoka, K. (1992). *Diagnostic Knowledge States in Algebra Using the Rule Space Model*. (Technical Report RR-92-57-ONR). Nueva Jersey: Educational Testing Service.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Embretson, S. E. (1999). Issues in the Measurement of Cognitive Abilities. In: S. E. Embretson & S. L. Hershberger (Eds.) *The New Rules of Measurement: What Every Psychologist and Educator*

- Should Know* (pp. 1-15). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gierl, M. J., & Cui, Y. (2008). Defining characteristics of diagnostic classification models and the problem of retrofitting in cognitive diagnostic assessment. *Measurement: Interdisciplinary Research and Perspectives*, 6, 263-268.
- Gierl, M. J., Leighton, J. P., & Hunka, S. M. (2007). Using the Attribute Hierarchy Method to Make Diagnostic Inferences About Examinees' Cognitive Skills. In: J. P. Leighton & M. J. Gierl (Eds.) *Cognitive Diagnostic Assessment: Theory and Applications* (pp. 242-274). Nueva York: Cambridge University Press.
- Gierl, M., Roberts, M., Brito, C., & Gotzmann, A. (2009, abril). *Using Judgments from Content Specialists to Develop Cognitive Models for Diagnostic Assessments*. Artículo presentado en Annual Meeting of the National Council on Measurement in Education (NCME), San Diego, CA.
- Gorin, J. (2006). Test design with cognition in mind. *Educational Measurement: Issues and Practice*, 25(4), 21-35.
- Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of Item Response Theory*. California: SAGE Publications.
- Im, S. (2007). *Statistical Consequences of Attribute Misspecification in the Rule Space Model*. (Unpublished doctoral dissertation). Columbia University, Nueva York, EE. UU.
- Leighton, J., & Gierl, M. (2007). Defining and Evaluating Models of Cognition Used in Educational Measurement to Make Inferences About Examinees' Thinking Processes. *Educational Measurement: Issues and Practice*, 26(2), 3-16.
- Leighton, J., Gierl, M., & Hunka, S. (2002, abril). *The attribute hierarchy model for cognitive assessment*. Louisiana: Artículo presentado en Annual Meeting of the National Council on Measurement in Education (NCME).
- Messick, S. (1984). The Psychology of Educational Measurement. *Journal of Educational Measurement*, 21(3), 215-237.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Mislevy, R. (2006). Cognitive Psychology and Educational Assessment. In: R.L. Brennan (Ed.), *Educational Measurement* (4 ed., pp. 257-305). Nueva York: American Council on Education/Macmillan.
- Mislevy, R., Steinberg, L., & Almond, R. (2003). On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3-62.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Nichols, P. (1994). A framework for developing cognitively diagnostic assessments. *Review of Educational Research*, 64, 575-603.
- Roussos, L. A., DiBello, L. V., Henson, R. A., Jang, E., & Templin, J. (2010). Skills Diagnosis for Education and Psychology With IRT-Based Parametric Latent Class Models. En S. E. Embretson (Ed.) *Measuring Psychological Constructs: Advances in Model-Based Approaches*. Washington: American Psychological Association.
- Rupp, A., & Mislevy, R. (2007). Cognitive Foundations of Structured Item Response Models. In: J. P. Leighton & M. J. Gierl (Eds.) *Cognitive Diagnostic Assessment: Theory and Applications* (pp. 205-241). Nueva York: Cambridge University Press.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. Nueva York: The Guilford Press.
- Shaw, D. S. (1986). *Effects of adaptive diagnostic testing on two types of computerized remediation*. (Unpublished doctoral dissertation), University of Illinois at Urbana-Champaign.

- Snow, R.E., & Lohman, D.F. (1989). Implications of cognitive psychology for educational measurement. In R.L. Linn (Ed.), *Educational Measurement* (3 ed., pp. 263-331). Nueva York: American Council on Education/Macmillan.
- Sternberg, R. (1986). *Las capacidades humanas: un enfoque desde el procesamiento de la información*. Barcelona: Editorial Labor.
- Pellegrino, J., Baxter, G., & Glaser, R. (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.
- Pozo, J. I. (2002). *Teorías cognitivas del aprendizaje*. Madrid: Ediciones Morata.
- 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. (1987). Validation of cognitive sensitivity for item response curves. *Journal of Educational Measurement*, 24(3), 233-245.
- 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). Nueva Jersey: Erlbaum.
- Tatsuoka, K. (2009). *Cognitive Assessment: An Introduction to the Rule Space Method*. Nueva York: Routledge Taylor & Francis Group.

Recibido: 20 de mayo de 2015
Aceptado: 16 de setiembre de 2015