

Utilización del modelo G-DINA para promover la evaluación como forma de aprendizaje

Álvaro Artavia Medrano*

Universidad de Costa Rica, Facultad de Educación, Costa Rica.

Recibido: 19 octubre 2015

Aceptado: 14 noviembre de 2015

RESUMEN. Los modelos de diagnóstico cognitivo se han considerado como un enfoque interdisciplinario en el que se unen la psicología cognitiva con el análisis estadístico. Mediante la utilización del modelo G-DINA, se realizó un análisis de ajuste posterior a un subtest a partir de la Prueba Nacional de Bachillerato en Matemática aplicada en el año 2006 a estudiantes costarricenses, la cual constituye un requisito para egresar de la educación secundaria. Con una estructura de 10 atributos, se explica el rendimiento en los 37 ítems correspondientes a álgebra y funciones elementales. Los resultados muestran que la verificación de proposiciones es el atributo de mayor dificultad, mientras que la factorización de polinomios es el más fácil. Asimismo, se presentan perfiles de dominio de atributos y ajuste del modelo propuesto.

PALABRAS CLAVE. Evaluación Diagnóstica Cognitiva, Modelo G-DINA, Aprendizaje.

Using the G-DINA model to promote assessment for learning

ABSTRACT. Cognitive diagnosis models have been considered as an interdisciplinary approach to linking cognitive psychology and statistical analysis. Using the G-DINA model, a retrofitting analysis was performed on a subtest from the National Mathematics Test from 2006 given to Costa Rican high school students. The success in the 37 items related to algebra and elementary functions was explained by 10 cognitive attributes. Results show that verification of statements is the most difficult attribute while factorization is the easiest. Also, attribute mastery profiles and model fit are presented.

KEY WORDS. Cognitive Diagnostic Assessment, G-DINA Model, Learning.

1. INTRODUCCIÓN

La evaluación diagnóstica cognitiva se considera como un desarrollo de la medición psicoeducativa que permite tener información más detallada acerca del dominio de habilidades específicas

* Correspondencia: **Álvaro Artavia Medrano**. Dirección: Guadalupe, Goicoechea. Del antiguo Cine Reina, 225 metros al sur. Condominio Don Fernando, casa 24. San José, Costa Rica. Correo electrónico: alvaro.artavia@ucr.ac.cr

en la resolución de diversas situaciones evidenciadas en los ítems de pruebas en dicho ámbito, con mayor granularidad de la que permiten los modelos unidimensionales de teoría de respuesta a los ítems (TRI), tales como el modelo de Rasch.

Los métodos psicométricos tradicionales asumen o requieren la unidimensionalidad como una condición para la calibración de ítems y la estimación de parámetros. Por el contrario, los modelos de diagnóstico cognitivo asignan perfiles multidimensionales de habilidades al clasificar a las personas en cuanto a su dominio o no en cada una de las habilidades involucradas en una prueba determinada.

En consecuencia, el propósito fundamental de los modelos de diagnóstico cognitivo es el de “investigar la relación entre los procesos psicológicos y las estrategias que subyacen al rendimiento en ítems de una prueba determinada y las respuestas dadas a tales ítems mediante análisis estadísticos sofisticados” (Ravand y Robitzsch, 2015, p. 1).

Por ello, a diferencia de ubicar a las personas con una puntuación a lo largo de una escala continua, como es el caso de la TRI, los modelos de diagnóstico cognitivo dan una explicación a las respuestas a los ítems a partir de evidencia empírica acerca de los procedimientos y las estrategias empleadas en la resolución de los ítems que conforman una prueba, mediante un conjunto de atributos cognitivos claramente definidos pero que a la vez se interrelacionan.

De acuerdo con Artavia (2013)

“los modelos cognitivos permiten interpretaciones de los resultados en pruebas de rendimiento y guían sobre los aspectos que deben ser reforzados en siguientes etapas del aprendizaje escolar. De igual manera, permiten enfatizar en el desarrollo de habilidades complejas y también atender concepciones equivocadas que el estudiantado muestre y que les dificulte su progreso hacia niveles superiores de aprendizaje” (p. 179).

La vinculación entre la psicometría y la psicología cognitiva para identificar la forma en que las personas organizan su conocimiento, diagnosticar las concepciones erróneas que evidencien y proponer acciones para el mejoramiento del aprendizaje, se ha estudiado en diversas investigaciones y propuestas, por ejemplo, Tatsuoka (1983), Messick (1984), Nichols (1994), Pellegrino, Baxter y Glaser (1999), Leighton, Gierl y Hunka (2002), Gorin (2006), Jang (2008), Rupp, Tempelin y Henson (2010) y de la Torre (2013).

Los modelos de diagnóstico cognitivo se basan en la adquisición de conocimiento a partir de un procesamiento sistemático de la información, en el que las personas aprenden conceptos mediante el razonamiento y emplean estrategias cognitivas y metacognitivas que les permiten transferir nuevos conocimientos a otras tareas (Jang, 2008). Esta perspectiva permite enfocarse en el papel del conocimiento, su contenido y los procesos cognitivos necesarios para la resolución de problemas, esto es, la pericia se puede descomponer en un conjunto de estructuras de conocimiento que se aprenden, lo cual tiene implicaciones en el desarrollo de ambientes de aprendizaje para facilitar su adquisición (Nokes, Schunn y Chi, 2010).

A lo largo de su aún reciente desarrollo en la literatura técnica, los modelos de diagnóstico cognitivo se han empleado en dos formas diferentes, ya sea en el análisis de pruebas existentes, o bien, en el diseño de conjuntos de ítems para propósitos diagnósticos. En el campo de la evaluación de los aprendizajes, la mayoría de las aplicaciones de estos modelos se basa en la primera forma de utilización, conocida como análisis post hoc o de ajuste posterior [retrofitting], como un intento por disponer de información que permita el diagnóstico de las habilidades empleadas en

la resolución de los ítems y que subyacen a su construcción, aun cuando el diseño original de la prueba no lo hubiese considerado así (Roussos, Dibello, Henson, Jang y Templin, 2010). Por ser el método más comúnmente empleado y por el interés del estudio que se presenta en este artículo, se describirán a continuación los pasos para la realización de análisis de ajuste posterior, según Ravand y Robitzsch (2015).

El primer paso consiste en especificar los atributos requeridos para desempeñarse exitosamente en una prueba determinada, lo cual se lleva a cabo mediante teorías propias del dominio específico de conocimiento, revisión de literatura, protocolos verbales y protocolos de especialistas, así como análisis de contenido para los ítems. En este estado se pretende tener evidencias de la representación del constructo, según la propuesta de Embretson (1983). En este contexto, 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. De esta manera, “aunque un atributo no es una estrategia, brinda los cimientos para las estrategias. Además, el conjunto de atributos organizados en una estrategia cumplen el papel momentáneo de resolver el problema, pero no necesariamente permanecen agrupados como una estrategia” (Leighton, Gierl y Hunka, 2002, p. 4).

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

Aunque es posible descomponer atributos complejos en tareas más simples o específicas, esto también aumenta la cantidad de atributos para los ítems, lo cual puede hacer imposible la estimación de probabilidades de dominio de atributos en un modelo de diagnóstico cognitivo. Por ello, lo recomendable es trabajar una cantidad de atributos que sea estadísticamente manejable de acuerdo con la cantidad de ítems involucrados y el tamaño de la muestra de personas que contestan la prueba.

El segundo paso en el análisis de pruebas ya existentes consiste en la elaboración de una matriz binaria en la que se especifiquen las relaciones entre atributos y los ítems que conforman una prueba; a esta matriz se le conoce como matriz de incidencia o matriz Q (Tatsuoka, 1983, 1990, 2009) y su validación teórica y empírica es lo que concede el poder de diagnóstico a los modelos basados en ella.

El tercer paso consiste en la especificación del modelo por emplear, el cual especifica las relaciones entre los atributos postulados como estructura subyacente a la resolución de los ítems de una prueba. En particular, estas relaciones pueden ser de naturaleza conjuntiva, compensatoria o general.

Finalmente, el cuarto paso consiste en estimar los perfiles de dominio de atributos, a partir de las respuestas observadas, esto es, poder estimar la probabilidad de dominio de los atributos que subyacen a la estructura cognitiva de una prueba, tomando el desempeño real de cada persona en dicha prueba.

2. EVALUACIÓN DEL DESEMPEÑO EN ÍTEMS DE MATEMÁTICA

Según Onrubia, Rochera y Barberà (2001), la Matemática se ha constituido en uno de los domin-

ios específicos más estudiados por la investigación psicoeducativa, fundamentalmente por dos razones: 1) las relaciones entre los procesos de resolución de problemas, los lenguajes formales y los sistemas notacionales de representación que median el aprendizaje escolar, y 2) las dificultades que muchas personas muestran para aprender Matemática en su etapa escolarizada, lo cual se ha reconocido habitualmente y documentado de manera amplia.

De acuerdo con Schoenfeld (2007), se ha originado un cambio en la enseñanza y el aprendizaje de la Matemática: pasar de un énfasis exclusivo en el conocimiento, es decir, lo que saben –o deberían saber– las personas, a enfatizar en qué pueden hacer con ese conocimiento, es decir, la idea es que no es suficiente tener el conocimiento, sino saber utilizarlo en las circunstancias apropiadas, lo cual es un componente esencial de la capacidad matemática.

La resolución de problemas matemáticos puede subdividirse en dos partes principales: la representación del problema, que es la conversión de un problema verbal en una representación interna y la solución del problema, que es la aplicación de los operadores matemáticos legales a la representación interna para poder llegar a una respuesta final. La solución del problema implica el desarrollo y el control de un plan de solución y su correspondiente ejecución.

Según Mayer (1986), algunos conocimientos que pueden ser relevantes en la representación y la resolución de problemas son: 1) conocimiento lingüístico, 2) conocimientos generales, 3) conocimiento esquemático, 4) conocimientos estratégicos, y 5) conocimientos algorítmicos. Los dos primeros son necesarios en la traducción del problema, el tercero se requiere en la integración del problema; el cuarto, en la planificación de la solución y, finalmente, el quinto se necesita en la ejecución de las soluciones. A continuación se describirá su importancia en la resolución de problemas.

El primer paso en la representación del problema es el de ser capaz de traducir cada proposición del problema en una representación interna, para lo que se requiere de algún conocimiento del lenguaje y de la realidad objetiva. Las personas pueden diferir en su capacidad para traducir correctamente las proposiciones y tales diferencias pueden estar relacionadas con los conocimientos lingüísticos y generales.

El siguiente componente en la resolución de problemas matemáticos es el de agrupar las proposiciones textuales del problema en una representación coherente, para lo que se debe poseer algún conocimiento del tipo de problemas que se pueden presentar (conocimiento esquemático). Una persona que intente relacionar este problema con su experiencia anterior en problemas similares podrá darse cuenta de que las proposiciones son contradictorias. Particularmente, la comprensión de los problemas presentados mediante palabras está influida por el hecho de si las personas tienen un esquema apropiado para el problema o pueden acceder a él.

El tercer nivel corresponde a la planificación de la solución o diseño de un plan de solución, para lo que se requiere que la persona posea y utilice su conocimiento estratégico.

Finalmente, la ejecución de la solución requiere que la persona sea capaz de efectuar operaciones, para lo que necesita algún conocimiento de los procedimientos de solución, es decir, conocimientos algorítmicos. No obstante, una persona que utilice un algoritmo con una o más fallas puede ser capaz de generar la respuesta correcta en algunos problemas, pero otras veces cometerá errores. Ciertamente en estos casos se pone de manifiesto la importancia de analizar el espacio del problema, pues con él se identifican las reglas o las estrategias empleadas –y que están en el espacio de reglas– al enfrentar la tarea y que explican la naturaleza de las respuestas que la persona proporciona en una situación particular.

Ahora bien, dado que cada persona adquiere, procesa y organiza el conocimiento de forma diferente, se requiere una comprensión más profunda de cómo las personas llevan a cabo tareas cognitivas, pues de esa manera será mayor la capacidad para evaluar el desempeño del estudiantado y con ello proponer acciones para orientar el mejoramiento del aprendizaje, una de las principales razones de ser de los modelos de diagnóstico cognitivo.

3. EL MODELO G-DINA

En el ámbito educativo, los modelos de diagnóstico cognitivo brindan información específica que permite evaluar las fortalezas y debilidades del estudiantado en un dominio específico de conocimiento.

Aunque existen varios modelos para llevar a cabo estos análisis, uno de los más simples y, por tanto, menos restrictivos en sus condiciones, es el modelo G-DINA (de la Torre, 2011, 2013), cuya base fundamental radica en el modelo DINA [deterministic input noisy and gate model] propuesto por de la Torre y Douglas (2004). Al igual que otros, el modelo G-DINA se basa en una matriz de incidencia (o matriz Q) en la que se establecen las relaciones existentes entre ítems y los atributos que subyacen a su resolución.

El modelo DINA divide a las personas en dos clases para cada ítem: aquellas que dominan todos los atributos requeridos para tal ítem y aquellas que no los dominan, por lo que modela la probabilidad de acertar un ítem a partir de dos parámetros: 1) la posibilidad de que una persona se equivoque al responder un ítem aun cuando domina todos los atributos requeridos (denota el parámetro de “equivocaciones”), y 2) la probabilidad de que una persona acierte un ítem a pesar de que no posea todos los atributos que se requieren (denota el parámetro de “adivinación”). La probabilidad de respuesta a un ítem se escribe así:

$$P(X_{ij} = 1 | \xi_{ij}, s_j, g_j) = (1 - s_j)^{\xi_{ij}} g_j^{(1 - \xi_{ij})}$$
 donde X_{ij} es la respuesta de la persona i al ítem j .

Una generalización del DINA la constituye el modelo G-DINA [generalized deterministic inputs, noisy “and” gate], el cual propone que al aplicar las restricciones apropiadas, muchos de los modelos de diagnóstico cognitivo resultan casos particulares del modelo general. Específicamente, a partir de un conjunto de atributos correspondientes a un ítem, el modelo G-DINA brinda una probabilidad de éxito para cada una de las posibles combinaciones de dominio de atributos, incluyendo no dominar ninguno de los atributos requeridos en la resolución del ítem, según lo establecido en la matriz de incidencia. Por ello, el modelo G-DINA permite examinar la contribución específica de un atributo en el desempeño en un ítem determinado, así como el efecto de la interacción que se produce al dominar varios atributos al mismo tiempo (de la Torre, 2013).

4. MÉTODO

Los datos provienen de la Prueba de Bachillerato en Matemática aplicada en el año 2006 a estudiantes costarricenses, de manera censal, puesto que es un requisito para egresar de la educación secundaria. Para los propósitos de este estudio, se utilizó un muestreo aleatorio simple y se calculó el tamaño de muestra para estimar una proporción con el 95% de confianza, con lo que se obtuvo $n = 384$. Para cada persona, se contó con 37 variables puntuadas de manera dicotómica (1 en caso de acierto, 0 en caso contrario).

Para obtener la puntuación total en la prueba para cada persona, se sumaron las puntuaciones de las 37 variables mencionadas que corresponden a los ítems de interés para el estudio, los cuales corresponden a los temas de álgebra y funciones. Específicamente, en el primer caso, los ítems midieron factorización de polinomios, expresiones algebraicas fraccionarias, ecuaciones cuadráticas, sistemas de ecuaciones lineales y problemas de aplicación. En el segundo caso, los ítems midieron conceptos generales de funciones, función lineal, rectas paralelas, rectas perpendiculares, intersección de rectas, función cuadrática, clasificación basada en el codominio, función inversa, problemas de aplicación y conceptos generales de función exponencial y función logarítmica.

En el estudio cuyos resultados aquí se presentan, se consideró de particular interés poder investigar el desempeño del estudiantado en los componentes de álgebra y funciones debido al historial de rendimiento que se ha evidenciado en los últimos años en dichas pruebas: una elevada proporción de estudiantes que acierta los ítems de álgebra y lo contrario sucede en el componente de funciones. La utilización de calculadora científica en las pruebas, así como el bajo dominio de conceptos asociados con funciones se han discutido constantemente en diversos escenarios nacionales como explicación de tales diferencias.

5. RESULTADOS

5.1 Propiedades psicométricas

La puntuación total en la prueba para cada persona se obtuvo al sumar las puntuaciones de los 37 ítems considerados. La media aritmética de la puntuación total es de 23,09 con una desviación estándar de 5,87; las puntuaciones van desde 5 hasta 37 puntos.

Debido a que estos ítems fueron seleccionados con un propósito definido, pero son parte de una prueba compuesta por 60 ítems (se excluyeron los correspondientes a geometría y trigonometría), se hicieron análisis psicométricos para determinar si era posible considerar este conjunto de ítems como un subtest. Para ello, se calculó el alfa de Cronbach y se obtuvo 0,82 el cual se puede considerar como indicador de alta consistencia interna; este valor se obtuvo con la ayuda del software IBM SPSS Statistics 22 (SPSS, 2013). Adicionalmente se realizaron análisis para reducir la escala, esto es, la posibilidad de eliminar ítems si con eso se contribuye a una mejor consistencia en los resultados de la prueba; no obstante, no se encontró un aporte que se considerara relevante en este sentido, por lo que se mantuvieron los 37 ítems.

En cuanto a las características del conjunto de ítems seleccionados, se calcularon los índices de dificultad y de discriminación según la teoría clásica de los test y el modelo de Rasch; ambos con la ayuda del software jMetrik 4.0 (Meyer, 2014). En el caso de la teoría clásica, la proporción de respuestas correctas varió de 0,18 a 0,93 con una media aritmética igual a 0,62 y una desviación estándar de 0,21. Para el modelo de Rasch, la aptitud varió de $-2,31$ a $2,52$ lógitos; todos los ítems se ajustan a dicho modelo, según indican las medidas de ajuste interno [*infit*] y externo [*outfit*].

5.2 Matriz de incidencia

La matriz Q empleada en este estudio es una adaptación de la propuesta por Artavia (2014). En dicha investigación, el autor propuso un modelo cognitivo formado por 18 atributos para 55 ítems; no obstante, al extraer un subconjunto de tales ítems con propósitos investigativos, dicha matriz fue adaptada en sus dimensiones y se finalizó en 10 atributos para 37 ítems. En la tabla 1 se indican cuáles son los atributos empleados para estos análisis.

Tabla 1. Atributos cognitivos de la prueba

Atributo	Descripción
AT1	Utilización de métodos de factorización
AT2	Revisión de opciones de un ítem
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	Distinción entre variables y parámetros
AT10	Utilización de propiedades de potencias o logaritmos

Con el propósito de recabar evidencias de validez para la matriz Q propuesta, se llevó a cabo un análisis de regresión múltiple con el valor de χ^2 (calculado con el modelo de Rasch) como variable dependiente y los vectores binarios de atributos como variables independientes. Se obtuvo un coeficiente de determinación de 0,89 y un coeficiente ajustado de 0,85; esto es, más del 85% de la varianza en la dificultad de los ítems se explica por los atributos involucrados en ellos.

Tabla 2. Atributo y dificultad

Atributo	Dificultad
AT1	0,79
AT2	0,72
AT3	0,56
AT4	0,64
AT5	0,59
AT6	0,35
AT7	0,32
AT8	0,31
AT9	0,48
AT10	0,71

Como se muestra en la tabla 2, el AT1 (utilización de métodos de factorización) fue el más fácil, dado que es dominado por el 79% de las personas examinadas; a dicho atributo le siguen el 2 y el 10, que corresponden a la revisión de opciones de un ítem (72% de dominio) y la utilización de propiedades de potencias o logaritmos (71%). Los dos atributos que se consideran de mayor dificultad son el 8 y el 7 que corresponden a la verificación de proposiciones (dominado por un 31%) y la determinación o el cálculo de conceptos básicos de funciones (32% de dominio).

Las tablas 3 y 4 contienen ejemplos de parámetros para los ítems según el modelo G-DINA. En ellas, se despliega el número de ítem en la primera columna; se detallan los atributos requeridos en su resolución de acuerdo con lo indicado en la matriz Q, en la segunda columna; las posibles combinaciones entre los atributos requeridos o bien, los patrones de dominio de atributos, en la

tercera columna y, finalmente, en la cuarta columna aparece la probabilidad de acertar el ítem según el dominio que se tenga de los atributos involucrados en su resolución.

Tabla 3. Parámetros de tres ítems de álgebra

Ítem	Atributos requeridos	Combinaciones de atributos	Probabilidad
1	AT1 y AT2	00	0,89
		10	0,67
		01	1,00
		11	0,99
2	AT1, AT2 y AT3	000	0,23
		100	0,33
		010	0,00
		001	0,84
		110	1,00
		101	0,87
		011	0,67
		111	0,83
3	AT1, AT2 y AT3	000	0,47
		100	0,48
		010	1,00
		001	1,00
		110	1,00
		101	0,72
		011	0,66
		111	0,95

Como se muestra en la tabla 3, aquellas personas que no dominan ninguno de los atributos requeridos en el ítem 1, tienen una probabilidad de 0,89 de acertar dicho ítem. Por otra parte, quienes dominan solo el AT1 (utilización de métodos de factorización) tienen un 0,67 de probabilidad de acertar, mientras que quienes dominan solo el AT2 (revisión de opciones) tienen la certeza de éxito en este ítem; prácticamente la misma probabilidad en el caso de dominar ambos atributos.

Para los ítems 2 y 3, la situación es diferente a la ya descrita. No dominar ninguno de los atributos requeridos brinda probabilidades de acierto iguales a 0,23 y 0,47, respectivamente. En particular, para el ítem 2, las personas que dominan únicamente el AT2 tienen una probabilidad nula de acierto. No obstante, para el ítem 3, basta con que las personas dominen el AT2 o el AT3 (simplificación de expresiones) para garantizar el éxito en este ítem.

Según lo expuesto en la tabla 4, las personas que no dominan ninguno de los atributos requeridos para resolver el ítem 13, tienen un 0,14 de probabilidad de acertarlo, mientras que al dominar los dos atributos requeridos, su probabilidad es de 0,66. Por otra parte, para el ítem 33, cuya resolución está vinculada con el dominio de tres atributos, se tiene que un 0,39 de probabilidad de

acertarlo sin dominar ninguno de los atributos y si únicamente se domina la distinción entre variables y parámetros (AT9) se estima una probabilidad nula para el acierto; no obstante, se estima como suficiente el dominio de la revisión de opciones (AT2) y la interpretación de la información contenida en el enunciado (AT6), para tener la certeza de aprobar este ítem en particular.

Tabla 4. Parámetros de tres ítems de funciones

Ítem	Atributos requeridos	Combinaciones de atributos	Probabilidad
13	AT7 y AT8	00	0,14
		10	0,08
		01	0,00
		11	0,66
22	AT2, AT6 y AT7	000	0,31
		100	0,14
		010	0,35
		001	0,45
		110	0,37
		101	0,00
		011	0,00
		111	0,59
33	AT2, AT6 y AT9	000	0,39
		100	0,41
		010	0,51
		001	0,00
		110	1,00
		101	0,14
		011	0,27
		111	0,49

5.4 Fiabilidad en la clasificación

Cui, Gierl y Chang (2012) desarrollaron un método para determinar la fiabilidad en la clasificación de personas para un determinado modelo de diagnóstico cognitivo. Sus dos principales conceptos son la precisión y la consistencia en la clasificación, por lo que crearon un índice para cada uno de ellos.

La consistencia en la clasificación se refiere al grado de acuerdo en las clasificaciones en dos administraciones independientes de la prueba o dos formas paralelas de él. De esta manera, la probabilidad de clasificar consistentemente a una persona seleccionada al azar, en dos administraciones o dos formas de la prueba, se conoce como índice de consistencia en la clasificación de diagnóstico cognitivo y se denota P_c .

Por otra parte, la precisión en la clasificación se refiere al grado de acuerdo entre la clasificación de la persona en clases latentes según sus patrones observados de respuestas y las verdaderas clases latentes de la persona. En consecuencia, la probabilidad de clasificar con precisión a una

persona seleccionada al azar, tomando como base sus respuestas a los ítems de una prueba, se conoce como índice de precisión en la clasificación de diagnóstico cognitivo y se denota P_a .

En la tabla 5 se muestran los valores de P_a y P_c para cada uno de los atributos empleados en el estudio que se muestra en este artículo.

Tabla 5. Consistencia y precisión en la clasificación para cada atributo

Atributo	P_a	P_c
AT1	0,29	0,88
AT2	0,30	0,96
AT3	0,47	0,90
AT4	0,47	0,83
AT5	0,85	0,77
AT6	0,91	0,84
AT7	0,92	0,85
AT8	0,87	0,80
AT9	0,87	0,79
AT10	0,31	0,97

A partir de la información de la tabla 5, el índice de precisión tiene una media aritmética de 0,63 con una desviación estándar de 0,28 y el índice de consistencia tiene un promedio de 0,86 con una desviación estándar de 0,07. De acuerdo con los índices propuestos, los atributos que permiten una mayor consistencia en la clasificación son el AT10 y el AT2, mientras que los atributos que permiten una mayor precisión en la clasificación son el AT7 y el AT6. Es relevante notar que los atributos que evidencian una precisión muy baja en la clasificación son AT1, AT2 y AT10 y son los atributos considerados como los más fáciles, según la tabla 2.

En la tabla 6 se muestran las posibles combinaciones de atributos para resolver el ítem 10 de la prueba, el cual, según el modelo de Rasch, es el más difícil de la prueba. Asimismo, las especificaciones de la matriz Q indican que es el único ítem que requiere de cuatro atributos: revisión de opciones (AT2), simplificación de expresiones (AT3), traducción y formulación de expresiones para resolver un problema (AT5) e; en particular, este último atributo es uno de los que resultan más difíciles, según las probabilidades de dominio que se muestran en la tabla 2 y que fueron obtenidas con el modelo G-DINA.

Tabla 6. Parámetros del ítem más difícil de la prueba

Combinaciones de atributos	Probabilidad
0000	0,00
1000	0,10
0100	0,29
0010	0,00
0001	0,00
1100	0,00
1010	0,04

1001	0,00
0110	0,00
0101	0,00
0011	0,00
1110	0,11
1011	0,81
1101	0,00
0111	0,23
1111	0,52

5.5 Ajuste del modelo

Como prueba de significancia de ajuste absoluto del modelo, se obtuvo $\max X^2 = 11,15$ y además, se concluye que existe un ajuste entre los datos y el modelo G-DINA con que se ha trabajado. Este coeficiente fue propuesto por Chen, de la Torre y Zhang (2013).

6. CONSIDERACIONES FINALES

En el ámbito educativo se emplean modelos de diagnóstico cognitivo con el propósito de identificar la presencia o ausencia de diversas habilidades que se requieren para resolver los ítems de una prueba, incluso para distintos tipos de granularidad o especificidad en la información, según se pretenda informar o bien, el tipo de audiencia a la que se dirija un informe de evaluación de rendimiento escolar.

En este artículo se presentó una aplicación del modelo G-DINA para investigar el desempeño del estudiantado en los componentes de álgebra y funciones en una prueba estandarizada de altas consecuencias para el estudiantado en la asignatura de Matemática, a nivel de secundaria.

Los métodos tradicionales con que se suelen analizar estas pruebas, han mostrado reiteradamente que se aciertan en mayor medida los ítems de álgebra que los de funciones, pero no han logrado describir cuáles componentes específicos son los de mayor dominio ni la probabilidad de acierto según combinaciones determinadas de estos atributos.

Al establecer una matriz de incidencia para explicar el rendimiento en la prueba con base en patrones observados de respuesta, el modelo G-DINA permite concluir que la verificación de proposiciones (AT8), la determinación o el cálculo de conceptos básicos de funciones (AT7) y la interpretación de la información contenida en el enunciado (AT6), son los atributos de mayor dificultad. No obstante, la utilización de métodos de factorización (AT1), la revisión de las opciones de un ítem (AT2) y la utilización de propiedades de potencias o logaritmos (AT10) son los de mayor facilidad.

A partir de lo expuesto en el marco de referencia teórica acerca de la evaluación del desempeño en ítems de Matemática, la capacidad en este dominio específico de conocimiento no puede continuar enfatizando únicamente en lo que las personas deberían saber, sino más bien su utilización en situaciones determinadas en las que pongan de manifiesto diversas estrategias con las que evidencien sus logros de aprendizaje según lo que han adquirido, procesado y organizado en cuanto al conocimiento se refiere.

Tanto las dificultades asociadas a los atributos que conforman la matriz Q , como las interpretaciones que se han hecho en términos del dominio conjunto y la estimación de la probabilidad asociada a acertar un ítem, parecieran ser el reflejo de prácticas educativas basadas fundamentalmente en el manejo algorítmico, en vez de la utilización de conceptos y habilidades tales como la interpretación de enunciados y proposiciones.

Se aboga por un cambio en las aulas de Matemática para que se refuercen situaciones de aprendizaje en las que se integre el conocimiento lingüístico, el esquemático y el estratégico, ya que el marcado énfasis en el conocimiento algorítmico podría llevar a conclusiones erróneas en cuanto a la explicación que se pretende dar de la naturaleza del desempeño estudiantil en esta asignatura.

Otro de los aportes de este artículo es la muestra de la utilización del software R (2015) para llevar a cabo análisis basados en modelos de diagnóstico cognitivo, específicamente en el paquete CDM 4.4 (Robitzsch, Kiefer, George y Uenlue, 2015). Este software es realmente eficiente en el tiempo empleado en la estimación de parámetros de ítems y probabilidades, es gratuito y permite realizar estos análisis de manera conjunta, con mínimas limitaciones que cada vez se van superando, en vez de rutinas o algoritmos complejos o que requieren conocimientos de programación y que podrían limitar en alguna medida, la producción de investigaciones en educación que se basen en modelos de diagnóstico cognitivo.

Finalmente, se propone la necesidad de continuar investigando acerca cómo inciden las diferentes combinaciones de atributos en la obtención de las respuestas correctas a los ítems de una prueba, sobre todo con el fin de poder diseñar pruebas que tomen estas especificaciones como punto de partida y de esta manera, contar con pruebas elaboradas con el propósito de comprender el rendimiento escolar con la complejidad que los modelos de diagnóstico cognitivo permiten y promueven.

REFERENCIAS

- Artavia, A. (2013). Incorporación de modelos cognitivos en la evaluación de los aprendizajes: un enfoque emergente. *Revista de evaluación educativa*, 2 (2), 162-183.
- Artavia, A. (2014). *Evaluación cognitiva diagnóstica en Matemática: modelo elaborado con el método rule space para estudiantes costarricenses de undécimo año*. Disertación doctoral no publicada, Universidad de Costa Rica, San José, Costa Rica.
- Cui, Y., Gierl, M. J., y Chang, H. (2012). Estimating classification consistency and accuracy for cognitive diagnostic assessment. *Journal of Educational Measurement*, 49 (1), 19-38.
- Chen, J., De la Torre, J., y Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, 50(2), 123-140.
- De la Torre, J., y Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69 (3), 333-353.
- De la Torre, J. (2011). The Generalized DINA Model Framework. *Psychometrika*, 76 (2), 179-199.
- De la Torre, J. (2013). Application of the DINA Model Framework to Enhance Assessment and Learning. En M. Mo Ching Mok (Ed.), *Self-directed Learning Oriented in the Asia-Pacific, Education in the Asia-Pacific Region: Issues, Concerns and Prospects* 18. Springer Science+Business Media.

- Embretson, S. (1983). Construct validity: construct representation vs. nomothetic span. *Psychological Bulletin*, 93, 179-197.
- Gorin, J. (2006). Test design with cognition in mind. *Educational Measurement: Issues and Practice*, 25(4), 21-35.
- Jang, E. (2008). A framework for cognitive diagnostic assessment. En C. Chapelle, R. Chung y J. Xu (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 117-131). Ames, IA: Iowa State University.
- Leight, J., Gierl, M., y Hunka, S. (abril, 2002). *The attribute hierarchy model for cognitive assessment*. Louisiana: Artículo presentado en Annual Meeting of the National Council on Measurement in Education (NCME). Recuperado de <http://www.education.ualberta.ca/educ/psych/crame>
- Mayer, R. (1986). Capacidad matemática. En R. Sternberg (Ed.) *Las capacidades humanas: un enfoque desde el procesamiento de la información* (pp. 165-194). Barcelona: Editorial Labor.
- Messick, S. (1984). The Psychology of Educational Measurement. *Journal of Educational Measurement*, 21(3), 215-237.
- Meyer, P. (2014). *jMetrik 4.0* [Software de cómputo].
- Nichols, P. (1994). A framework for developing cognitively diagnostic assessments. *Review of Educational Research*, 64, 575-603.
- Nokes, T., Schunn, C., y Chi, M. (2010). Problem Solving and Human Expertise. En P. Peterson, E. Baker y B. McGaw (Eds.). *International Encyclopedia of Education*. (pp. 265-272). Oxford: Elsevier.
- Onrubia, J., Rochera, M., y Barberà, E. (2001). La enseñanza y el aprendizaje de las matemáticas: una perspectiva psicológica. En C. Coll, J. Palacios y A. Marchesi (Comps.). *Desarrollo psicológico y educación: Psicología de la educación escolar* (pp. 487-508). Madrid: Alianza Editorial.
- Pellegrino, J., Baxter, G., y 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.
- Ravand, H., y Robitzsch, A. (2015). Cognitive Diagnostic Modeling Using R. Practical Assessment. *Research & Evaluation*, 20(11), 1-12.
- Roussos, L., Dibello, L., Henson, R., Jang, E., y 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.
- Robitzsch, A., Kiefer, T., George, A., y Uenlue, A. (2015). *CDM 4.4*. Recuperado de <https://sites.google.com/site/alexanderrobitzsch/software>
- Rupp, A., Templin, J., y Henson, R. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. Nueva York: The Guilford Press.
- Schoenfeld, A. (2007). What is Mathematical Proficiency and How Can It Be Assessed? En A. Schoenfeld (Ed.) *Assessing Mathematical Proficiency*. (pp. 59-74). Nueva York: Cambridge University Press.

SPSS, Inc. (2013). *IBM SPSS Statistics 22* [Software de cómputo].

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. (1990). Toward an integration of item-response theory and cognitive error diagnosis. En N. Frederiksen, R. Glaser, A. Lesgold y 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.

THE R FOUNDATION FOR STATISTICAL COMPUTING. (2015). *R 3.2.1* [Software de cómputo].