UNIVERSIDAD DE COSTA RICA

FACULTAD DE EDUCACIÓN

INSTITUTO DE INVESTIGACIÓN EN EDUCACIÓN

ESCUELA DE FORMACIÓN DOCENTE

Informe final del proyecto 724-B1-276

Modelos diagnósticos de procesos, estrategias y estructuras de conocimiento de estudiantes costarricenses a partir de métodos psicométrico-cognitivos

Álvaro Artavia Medrano (investigador principal)

Alicia Gurdián Fernández (investigadora)

Agosto, 2014

Unidad de Información y Referencia
28 OCT. 2014
Instituto de Investigación en Educación INIE

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

003861

18.09.01
2186

Tabla de contenido

Información general 4
Información administrativa del proyecto 4
Resumen 4
Descriptores 5
Antecedentes 6
Introducción 6
Planteamiento del problema 9
Objetivo general 9
Objetivos específicos 10
Antecedentes del proceso de investigación 10
Referente teórico 18
Necesidad de un modelo cognitivo de resolución de problemas 18
La teoría del procesamiento de la información 19
Teoría del espacio del problema 22
Capacidad matemática 25
Competencias implicadas en la capacidad matemática 26
El método *rule space* 30
Representación de patrones de respuesta 32
Estados de conocimiento 34
Procedimiento metodológico 35
Tipo de investigación 35
Método empleado 36
Descripción de las técnicas utilizadas 37
Población a la que va dirigida la investigación 38
Selección de participantes 39
Descripción de procedimientos para recolección y análisis de la información 40

Análisis y discusión de los resultados 45
Características de los ítems y de la prueba 45
Distribución de las puntuaciones obtenidas 48
Análisis con el modelo de Rasch 49
Elaboración de la matriz de incidencia 51
Matriz de incidencia y su confiabilidad 53
Lista de atributos y su descripción 53
Matriz de incidencia o modelo de diagnóstico 58
Clasificación previa según el método *rule space* 59
Estadísticas descriptivas de las probabilidades de dominio de atributos 60
Conclusiones y recomendaciones 62
Bibliografía 66
Anexo 78

Información general

En este apartado se presentan aspectos que permiten conocer los diferentes componentes del proyecto, así como la información administrativa, el resumen de lo realizado y unas palabras clave o descriptores con los que será posible su ubicación en diferentes motores de búsqueda en bases de datos.

Información administrativa del proyecto

Número del proyecto:	724-B1-276
Nombre del proyecto:	Modelos diagnósticos de procesos, estrategias y estructuras de conocimiento de estudiantes costarricenses a partir de métodos psicométrico-cognitivos
Unidad base del investigador:	Escuela de Formación Docente, Facultad de Educación
Unidad de adscripción:	Instituto de Investigación en Educación
Programa al que pertenece:	Investigación para el Mejoramiento de los procesos de enseñanza y aprendizaje
Nombre de investigadores:	Álvaro Artavia Medrano (investigador principal) Alicia Gurdián Fernández (investigadora)
Vigencia del proyecto:	21/08/2011 al 31/05/2013 con prórroga al 31/08/2014

Resumen

El propósito fundamental de este proyecto de investigación consistió en desarrollar un modelo cognitivo que permita la formulación de inferencias acerca del desempeño en

una prueba de Matemática de estudiantes costarricenses de undécimo año. Para ello, se elaboró una matriz de incidencia o matriz Q en la que se explicitan las relaciones entre los ítems de la prueba y los atributos cognitivos correspondientes a su resolución, por lo que se considera el modelo cognitivo. Su validación se llevó a cabo mediante diversos análisis de regresión múltiple, así como por la tasa de clasificación de personas según el dominio del conjunto de habilidades y conocimientos que conforman el modelo.

Empíricamente, el proyecto de investigación mostró que las personas caracterizadas como de habilidades bajas tienden a no conglomerarse en estados de conocimiento, es decir, la explicación de sus patrones de respuesta se debe mayormente a errores aleatorios y sistemáticos de diversa índole. Este hallazgo, unido a una clara utilización de algoritmos en la resolución de los ítems de la prueba, son elementos que caracterizan el enfoque de enseñanza que ha predominado en Costa Rica.

Este proyecto favorece, entre otros, el establecimiento de políticas educativas innovadoras que vinculen los resultados de evaluaciones con las habilidades que se hacen imprescindibles para desarrollarse en la sociedad, el ámbito académico y laboral.

Descriptores

Evaluación diagnóstica, modelos de diagnóstico, métodos psicométrico-cognitivos, resolución de problemas

Antecedentes

En este segundo apartado se presenta el tema de investigación, su importancia y se reseña la problemática que se abordó en el proyecto, así como los objetivos por alcanzar.

Introducción

Uno de los propósitos de las pruebas en el ámbito educativo es brindar información acerca de los logros de aprendizaje del estudiantado. No obstante, los resultados de tales pruebas se presentan generalmente como la cantidad de aciertos o como una puntuación en una escala determinada, por lo que no ofrecen información acerca del conocimiento o las habilidades que las personas evidencian al responder correctamente un conjunto de ítems.

Actualmente, la mayoría de las pruebas en educación se basan en modelos psicométricos tradicionales, tales como la teoría clásica de los test (TCT) o la teoría de respuesta a los ítems (TRI), las cuales tienen una escasa conexión con los procesos, las estrategias y las estructuras de conocimiento involucradas en la resolución de ítems.

La falta de información cognitiva disponible al interpretar los resultados de pruebas de gran escala con modelos tradicionales ha sido una de las mayores limitantes para que se puedan utilizar como un medio de realimentar los procesos de enseñanza y aprendizaje; por el contrario, comúnmente tienen propósitos de certificación o selección, sin informar acerca de las fortalezas y las debilidades evidenciadas por el estudiantado.

Para poder satisfacer las crecientes demandas para los procesos de evaluación por parte de diversos sectores actuantes en el ámbito educativo, se han desarrollado modelos que reconocen los beneficios potenciales del análisis cognitivo de tareas, en los que las

Instituto de Investigación en Educación
Unidad de Información y Referencia

respuestas a los ítems siguen siendo importantes, pero no constituyen lo más relevante para comunicar como resultados de una evaluación.

El presente estudio promueve adoptar una concepción más integral de la evaluación de los aprendizajes, específicamente de la utilización de los resultados de pruebas estandarizadas de rendimiento académico, de tal manera que permita reportar información para docentes, estudiantes y diversos sectores involucrados en el sistema educativo nacional y que se consideren como un punto de partida para nuevos aprendizajes.

¿Qué justifica un cambio en el enfoque para la interpretación de puntuaciones de pruebas de gran escala en el ámbito educativo? En primer lugar, es realmente difícil comprender el rendimiento en una prueba sin considerar los procesos mentales subyacentes, sobre todo si lo que se desea es proponer intervenciones para mejorar el rendimiento de las personas con base en estudios científicos y técnicos, no en consideraciones propias de quienes se encargan de aspectos curriculares en dominios específicos o de lo que imaginan o desean desde una perspectiva personal y unitaria.

En segundo lugar, las explicaciones que tradicionalmente se han dado acerca del rendimiento de las personas fueron concebidas como argumentos deductivos, esto es la explicación se lograba a través de la derivación de observaciones a partir de leyes generales, sin consideraciones ontológicas más allá de las etiquetas observables. En respuesta a ello, se requiere el planteamiento de explicaciones causales del rendimiento en términos de lo que las personas comprenden, lo cual se puede hacer con la formulación de *modelos cognitivos*.

De esta manera, al enfatizar en procesos cognitivos y elaborar un modelo cognitivo de rendimiento, se enriquece el marco interpretativo para los resultados y las inferencias que se puedan hacer de tal rendimiento.

Dado que cuando se enseña en algún dominio específico no solo se desea que el estudiantado resuelva problemas y obtenga las respuestas correctas a los ítems de una prueba, sino comprender las razones del porqué ciertas respuestas son correctas y por qué algunos procedimientos son legítimos para llegar a ellas, no es posible hacerlo desde una posición externa a la persona. Por ello, los modelos cognitivos intentan brindar un punto de vista desde lo interior de la comprensión de esta, lo cual los hace importantes en la interpretación de las puntuaciones de pruebas.

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

Esta concepción de evaluación diagnóstica orientada a aspectos cognitivos también tiene gran incidencia en las prácticas de enseñanza, pues brinda conjuntos de conocimientos y habilidades que explican en alguna medida la psicología del desempeño en la resolución de situaciones en un dominio específico.

Planteamiento del problema

Uno de los mayores retos de las evaluaciones cognitivas diagnósticas es lograr una descripción detallada de los procesos involucrados y de las estructuras de conocimiento utilizadas por las personas para resolver los ítems de una prueba en particular. No obstante, la utilización de modelos cognitivos en la elaboración de diagnósticos se ha documentado escasamente en la literatura de la medición psicológica y educativa (Roberts, Brito, Gotzmann & Gierl, 2009).

Asimismo, se requiere contar con las evidencias psicológicas y estadísticas de que el modelo cognitivo de evaluación que se propone muestre que sus resultados sean válidos, lo que en este contexto implica medir los conocimientos y las habilidades que las personas utilizan en la resolución de problemas, pues de lo contrario no sería posible obtener la información diagnóstica que permita desarrollar y poner en práctica acciones educativas para mantener las fortalezas cognitivas del estudiantado y corregir los aspectos detectados como débiles.

Consecuentemente, el problema de investigación del presente estudio consiste en desarrollar un modelo cognitivo que permita la formulación de inferencias acerca del desempeño de las personas y brindar evidencias empíricas en cuanto a la adecuación o el ajuste del modelo cognitivo a un conjunto real de datos en una prueba del área de Matemática.

Objetivo general

Elaborar modelos de diagnóstico de procesos, estrategias y estructuras de conocimiento de estudiantes costarricenses a partir de métodos psicométrico-cognitivos.

Objetivos específicos

1. Identificar las habilidades cognitivas específicas, los procedimientos y los conocimientos (atributos cognitivos) que subyacen a una prueba de gran escala del sistema educativo nacional.

2. Clasificar las respuestas del estudiantado según patrones de dominio de habilidades, a partir de sus patrones de respuesta.

3. Elaborar procesos de divulgación que faciliten el acceso a la información de los alcances del proyecto.

Antecedentes del proceso de investigación

La arraigada tradición nominalista en la evaluación de los aprendizajes ha causado que las pruebas de gran escala no ofrezcan insumos de mayor utilidad para reflejar procesos cognitivos de niveles asociados al aprendizaje. Esto se evidencia incluso en las pruebas referidas a criterios, donde el rendimiento continúa siendo una propiedad de los ítems, no de las personas.

En respuesta a ello, diversas investigaciones realizadas por Calfee (1981), Tatsuoka (1983), Messick (1984), Bejar (1984), Frederiksen (1990), Nichols (1994), Pellegrino, Baxter y Glaser (1999), Junker y Sijtsma (2001), Leighton, Gierl y Hunka (2002), Mislevy, Steinberg y Almond (2003), Ferrara (2006), Gorin (2006), Rupp, Templin y Henson (2010), entre otros, han planteado la necesidad de combinar las ciencias cognitivas con la psicometría para identificar la forma en que las personas organizan su conocimiento, diagnosticar las

Instituto de Investigación en Educación
Unidad de Información y Referencia

concepciones erróneas que evidencien y con ello proponer acciones para el mejoramiento del aprendizaje.

En particular, Nichols (1994) utilizó el término *evaluación cognitiva diagnóstica*, con el propósito de resaltar la integración de la psicología cognitiva en el diseño y la interpretación de los resultados de evaluaciones, así como en la elaboración de diagnósticos sobre las características de las personas en cuanto a sus procesos cognitivos. Por su parte, Rupp (2007) considera que dicho término tiene algunas restricciones semánticas, pues no permite considerar tendencias desarrolladas recientemente para integrar la psicología cognitiva con la medición, pero que no tienen un propósito diagnóstico (por ejemplo, la generación automatizada de ítems). En consecuencia, propone los términos "psicometría cognitiva", "modelos psicométrico-cognitivos" e "investigación psicométrico-cognitiva".

Entre las propuestas pioneras de evaluaciones con orientaciones cognitivas, Brown y Burton (1977) desarrollaron modelos diagnósticos para errores procedimentales asociados a habilidades matemáticas básicas, con el propósito de explicar por qué una persona comete un error en vez de únicamente identificar en cuáles errores incurre.

Dada la cantidad de errores que pueden cometer las personas, así como sus combinaciones, los conjuntos de ítems para medirlos deben ser construidos con sumo cuidado incluso para poder diferenciarlos. Para ello, crearon un sistema de cómputo llamado DEBUGGY con la pretensión de construir un extenso catálogo de errores de procedimiento con una definición precisa al efectuar restas.

Aunque el trabajo de Brown y sus colegas mostró un nuevo enfoque para el diseño de instrumentos para la evaluación diagnóstica, investigaciones posteriores demostraron que un patrón de errores sistemáticos para una persona en particular no se mantiene estable en distintas pruebas. Por ejemplo, Tatsuoka, Birenbaum y Arnold (1989) concluyeron que una práctica común entre personas que no dominan las operaciones con números negativos

es no utilizar consistentemente las reglas para tales operaciones. En contraste, aquellas personas que utilizaban las reglas correctas, sí mostraban un patrón estable en la aplicación de tales reglas.

Tatsuoka (1983) utilizó la TRI para desarrollar el modelo *rule space*, con el que se podía predecir la probabilidad de la utilización de reglas erróneas al efectuar sumas y restas con números enteros.

La idea original del método consistió en representar la respuesta de una persona como un punto en un gráfico bidimensional, luego se graficaban también respuestas simuladas para cada regla errónea y cada puntuación verdadera obtenida en el conjunto de ítems. Si la representación gráfica de la respuesta de la persona quedaba superpuesta en la representación de una regla errónea, entonces se diagnosticaba que la persona incurría en ese error de procedimiento. Asimismo, también se podía tomar en cuenta qué tan cerca estaban las representaciones de las respuestas de las representaciones de los errores.

Las primeras aplicaciones del *rule space* enfatizaron sobre áreas o contenidos sumamente específicos de la Matemática, tales como operaciones con 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 respuestas.

Birenbaum, Kelly y Tatsuoka (1992, 1993) investigaron la estabilidad de los diagnósticos realizados a partir del análisis de errores y de los análisis hechos con el *rule space*. Para ello, elaboraron una prueba de 32 ítems para medir la resolución de ecuaciones lineales con una variable. Todos los ítems eran de respuesta abierta y se les solicitó al estudiantado que dejaran constancia escrita de todos los pasos utilizados en la resolución. Se contó con una muestra de 231 estudiantes que cursaban octavo y noveno años (cuyas

Instituto de Investigación en Educación
Unidad de Información y Referencia

edades oscilaban entre 14 y 15 años) en una secundaria de Tel Aviv y fueron agrupados en alto y bajo rendimiento en Matemática.

Por otra parte, Birenbaum y Tatsuoka (1987), así como Birenbaum, Tatsuoka y Gutvirtz (1992) utilizaron el método para estudiar la influencia que tiene el formato de un ítem en los patrones de respuesta de las personas en una prueba.

Otro campo de aplicación del método *rule space* ha sido en pruebas de Inglés (Kasai, 1997; Scott, 1998). En esos estudios se demostró la utilidad del método para validar un conjunto de atributos hipotéticos subyacentes al desempeño en Inglés, así como para proporcionar información diagnóstica de modo individual para las personas que contestaron la prueba.

Sheehan, Tatsuoka y Lewis (1993) utilizaron el *rule space* para diagnosticar y clasificar los estados de conocimiento relacionados con el desempeño en una encuesta nacional de alfabetización.

El método *rule space* se ha aplicado a varias pruebas estandarizadas tales como Scholastic Aptitude Test, SAT, Prueba de Aptitud Académica, PAA, Puerto Rico (Guerrero, 2001), Test of English as a Foreign Language, TOEFL, (Scott, 1998), National Assessment of Educational Progress, NAEP, (Yepes-Baraya, Tatsuoka, Allen, O'Sullivan & Hui, 1998), Trends in International Mathematics and Science Study, TIMSS y TIMSS-R (Tatsuoka Corter & Tatsuoka, 2004; Tatsuoka, Guerrero, Corter, Tatsuoka, Yamada, Xin, et al., 2006).

El College Board ha utilizado el método para generar informes desde octubre 2001 para la prueba Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test, PSAT/NMSQT (Milewski & Baron, 2002).

Por su parte, Katz, Martinez, Sheehan y Tatsuoka (1993) utilizaron análisis de protocolos verbales para guiar la construcción de un modelo para diagnosticar

conocimientos en el campo de la arquitectura. El *rule space* les permitió clasificar a profesionales en ejercicio, practicantes y estudiantes de la carrera de arquitectura de acuerdo con los atributos cognitivos identificados, principalmente en lo que se refiere a propiedades de elementos estructurales, como otros relativos a la composición adecuada de elementos propios de arquitectura en una computadora.

Xin, Xu y Tatsuoka (2004) investigaron la relación de la calidad del personal docente con el rendimiento y las habilidades cognitivas del estudiantado. Este es un estudio considerado como exploratorio en el que se utiliza el método *rule space* para someterlo a diversas pruebas de solidez y estabilidad. Para su trabajo, las autoras se basaron en 27 habilidades cognitivas y atributos de conocimiento previamente definidos para una prueba en la que se examinaban contenidos matemáticos por medio de 164 ítems. Estos ítems fueron validados estadísticamente y con la utilización de verbalizaciones hechas por las personas, quienes eran estudiantes de octavo año de cuatro países (Estados Unidos, Corea, Japón y Holanda) y fueron elegidos con el criterio de que esos países tienen características y resultados similares en la prueba TIMSS realizada en 1999. Posteriormente se hizo un análisis de comparación multinivel para explicar la variabilidad del logro del estudiantado en cada país según las cualidades del personal docente.

Las puntuaciones de dichas pruebas fueron descompuestas en tres categorías de habilidades cognitivas utilizando el método *rule space,* lo cual permitió que las investigadoras concluyeran que las características del personal docente no tienen una relación positiva consistente con cualesquiera de los atributos cognitivos determinados para el estudio.

En el año 2004, Tatsuoka, Corter y Tatsuoka elaboraron una investigación en la que analizaron qué tan bien se desempeñan, en una prueba TIMSS de Matemática, grupos de estudiantes que provienen de 20 países tanto con culturas como ambientes educativos

diferentes entre sí y cómo reelaborar este desempeño en términos de niveles de logro de conocimiento, habilidades y atributos de procesamiento. Para tales propósitos, lo primero que desarrollaron fue una lista donde se identificaron los atributos de los ítems y de esta manera los clasificaron en atributos de contenido, atributos de proceso y atributos de habilidades según el tipo de ítem.

Con base en los resultados, los investigadores concluyeron que la investigación del procesamiento cognitivo a nivel micro, las habilidades de pensamiento matemático y el conocimiento pueden conducir a nuevos hallazgos y brindar nuevos insumos para la valoración de modelos educativos de habilidades.

Chen, Gorin, Thompson y Tatsuoka (2006) plantearon una investigación en la que validaron los atributos cognitivos que se requieren para resolver los ítems de Matemática de la prueba TIMSS de 1999, tomando para ello a estudiantes taiwaneses. Con base en el método *rule space,* estimaciones de los parámetros de los ítems, análisis de regresión múltiple y comparaciones entre las probabilidades de atributos de los cuatro cuadernillos de examen, lograron el objetivo del estudio. En ese sentido, se considera que se logró establecer la conveniencia de un modelo cognitivo para representar el desempeño en la prueba TIMSS-1999 de estudiantes de octavo año de Taiwán, con lo que se hace un valioso aporte a la generalización de los constructos medidos por los ítems de la prueba de Matemática de TIMSS. Los investigadores consideraron de suma relevancia lo que denominan como validez de generalizabilidad y propusieron la verificación posterior de dos atributos más como parte de estudios futuros.

Dogan y Tatsuoka (2008) llevaron a cabo un estudio comparativo del desempeño de estudiantes de Turquía y Estados Unidos en la prueba TIMSS-R. Como parte de los últimos análisis con el método *rule space,* exploraron los estados de conocimiento más frecuentes y las trayectorias de aprendizaje derivadas a partir de ellos.

Instituto de Investigación en Educación
Unidad de Información
Herencia
INIE

En las aplicaciones más recientes del *rule space*, se encuentran el diagnóstico de habilidades para probar hipótesis estadísticas (Im & Yin, 2009), el cual es pionero en la utilización del método en esa área de conocimiento. Asimismo, Svetina, Gorin y Tatsuoka (2011) desarrollaron un modelo cognitivo con el *rule space* para la descripción de conocimientos, habilidades y aptitudes en una prueba de lectura crítica.

De esta manera, dado que la metodología propia del *rule space* ha tenido uno de los periodos más largos de desarrollo y aplicación en cuanto a estrategias de análisis, sus resultados han demostrado ser de gran utilidad en investigaciones en el campo educativo y psicológico (Tatsuoka, 2009).

De hecho, se encuentran varias disertaciones doctorales en las que se ha puesto de relieve dicha metodología. Guerrero (2001) investigó la equivalencia de constructo de la sección de Matemáticas de Scholastic Aptitud Test I (SAT I-M) y la Prueba de Aptitud Académica (PAA-M). Su propósito fue examinar si el método *rule space* resultaba ser una herramienta más efectiva que los métodos psicométricos tradicionales, específicamente para determinar si dos pruebas que siguen el mismo constructo y adaptadas a diferentes idiomas requieren que la misma persona utilice habilidades cognitivas similares o no al responder los ítems de ambas pruebas.

Por su parte, Dean (2006) enfocó su trabajo a la validación de un conjunto de atributos que podrían explicar el desempeño del estudiantado en la prueba TIMSS 95, específicamente en el campo de razonamiento matemático. Una vez que los validó, pudo clasificar exitosamente a más del 99% de las personas en estados de conocimiento y brindar información diagnóstica de gran relevancia en términos del dominio de los atributos identificados como factores explicativos del desempeño.

Chen (2006) aplicó la metodología *rule space* para valorar el desempeño de estudiantes taiwaneses en la prueba TIMSS 99-R y pudo determinar sus fortalezas y

debilidades a partir de los patrones de dominio de atributos. Además, realizó comparaciones grupales en cuanto a género y zona de ubicación del centro educativo en términos de las habilidades cognitivas en Matemática. Por su parte, Dogan (2006) investigó, con el *rule space*, la validez de constructo de la prueba de Matemática correspondiente al examen de admisión para la Universidad en Turquía.

Otros estudios doctorales han versado sobre aspectos teóricos del método *rule space*. Sung-Hoon (1990) investigó la clasificación de los patrones de respuesta a los ítems en grupos de concepciones erróneas desde un punto de vista estadístico de ajuste al método. Gierl (1996) evaluó los supuestos del ámbito cognitivo en que se basa el *rule space*, para lo que utilizó la prueba SAT aplicada en 1994. Finalmente, Im (2007) determinó las consecuencias estadísticas de una especificación incorrecta de los atributos cognitivos identificados y cómo esto puede afectar la inclusión o exclusión de ciertas habilidades en la estimación de la probabilidad de dominio de atributos según el método.

Actualmente se encuentra en desarrollo una evaluación para niños y niñas entre los 3 y los 7 años con el fin de brindar información al personal docente acerca del desarrollo de habilidades matemáticas tales como conteo, secuenciación, sumas y restas, entre otros. Se espera que el método *rule space* permita generar una evaluación cognitiva diagnóstica adaptada a esa población, con el fin de tener evidencias del conocimiento matemático y de sus posibilidades de desarrollo en cuanto a aprendizaje se refiere (Clements, Sarama, Tatsuoka & Tatsuoka, 2013).

Referente teórico

En este capítulo se plantea la base teórica que sustenta el proyecto de investigación, por lo que permite conocer la construcción teórica desde la que se abordó el estudio.

Necesidad de un modelo cognitivo de resolución de problemas

Como anteriormente se indicó, la mayoría de las teorías psicométricas están basadas en rasgos latentes, esto es, constructos hipotéticos que tienen la finalidad de describir las fuentes subyacentes de las diferencias individuales observadas en las puntuaciones de una prueba. Para lograrlo, afirman que un cierto número estable de características subyacentes o rasgos pueden explicar el rendimiento en una prueba.

No obstante, la necesidad de informarse con mayor profundidad acerca de los procesos que subyacen a la conducta inteligente provocó cambios en la investigación psicológica, puesto que las teorías psicométricas prácticamente no decían nada respecto a cuáles podrían ser tales procesos (Sternberg, 1986).

Así, más que identificar un factor del razonamiento humano, la psicología pretendía entender los procesos, las estrategias y las representaciones mentales que las personas utilizan en la ejecución de tareas de diversa índole. Dichas tareas son situaciones en las que se intenta alcanzar un objetivo y se hace necesario encontrar un medio para conseguirlo, por lo que se conocen como *problemas* (Chi & Glaser, 1986). Autores como Smith y Kosslyn (2008) conciben "problema" como una situación en la que no existe un camino inmediato aparente, habitual o rutinario, para alcanzar una meta, por lo que resolver un problema requiere sortear obstáculos para alcanzar un objetivo. Para ellos, "la meta final de la investigación sobre resolución de problemas ha sido identificar las estrategias que

utilizamos cuando nos enfrentamos a una situación nueva y debemos decidir una línea de acción" (p. 435).

La resolución de problemas es una aptitud cognitiva que caracteriza una de las actividades humanas más inteligentes. Desde la niñez, las personas resuelven activamente los problemas que se les presentan en la vida cotidiana. Las diferencias que se presentan en la capacidad de resolver problemas se basan en procesos cognitivos y organizaciones mentales que poseen (Chi & Glaser, 1986).

En lo que respecta al área educativa, con el propósito de modelar el desempeño cognitivo del estudiantado y brindar información específica acerca de las fortalezas y las debilidades en la resolución de problemas, se hace necesario un enfoque teórico que permita una comprensión profunda de las habilidades cognitivas que están vinculadas con el desempeño en una prueba en particular. En la siguiente sección se presentará el marco que fundamenta el trabajo en este campo.

La teoría del procesamiento de la información

Newell y Simon desarrollaron a partir de 1950 la *teoría del procesamiento de la información,* con lo que proporcionaron el marco explicativo más desarrollado y mayoritariamente aceptado para la resolución de problemas y cuyos supuestos fundamentales se mantienen luego de varias décadas de su formulación inicial (Eysenck & Keane, 2005; Holyoak, 1995; Klahr, 1989, 2000; Lohmann, 2000; vanLehn, 1989).

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.

Según estas teorías, una persona se enfrenta a un problema cuando debe resolver una tarea o situación determinada, pero sin conocer previamente la secuencia de acciones que debe aplicar para resolverla; tales acciones pueden ser físicas, perceptivas y exclusivamente mentales (Newell & Simon, 1972). Sin embargo, tener un problema implica al menos que se le dé cierta información inicial a la persona –lo que se desea, las condiciones, las herramientas y las operaciones– y el acceso a ciertos recursos.

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).

El posicionamiento teórico fundamental para el presente trabajo se basa en la perspectiva del procesamiento de información. En primer lugar, dado su amplio desarrollo e influencia en el estudio de la inteligencia humana por varias décadas, así como por el énfasis en los procesos cognitivos que componen la resolución de problemas en dominios específicos de conocimiento –propios del campo educativo– en cuyo contexto no se considera un requerimiento esencial llevar a cabo un análisis de las áreas del cerebro involucradas en los procesos mentales, sino más bien analizar los procesos propiamente

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

dichos (Rupp, 2007). 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 & Chi, 2010).

En segundo lugar, casi la totalidad de los trabajos desarrollados en el área de la medición educativa se basan en modelos psicométricos tradicionales (TCT o TRI). Por su parte, el procesamiento de la información es el marco que 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 se pueden desarrollar mejor conjuntamente en vez de considerarlas como entidades separadas (Sternberg, 1986).

En tercer lugar, el procesamiento de la información asume la importancia en la actuación inteligente tanto del conocimiento específico vinculado a dominios concretos, como de las estrategias generales, aplicables a una amplia variedad de dominios distintos.

Esto refleja un cambio en la concepción de las relaciones entre aprendizaje e inteligencia, al pasar de una concepción de inteligencia como condicionante causal del aprendizaje y del rendimiento escolar a la idea de que la inteligencia tiene un papel en el aprendizaje y también que el aprendizaje juega un papel decisivo en el desarrollo y el mejoramiento de la inteligencia, la cual puede optimizarse gracias a determinadas formas de intervención educativa (Coll & Onrubia, 2001).

Finalmente, la teoría del procesamiento de la información se encuentra en la tradición misma del desarrollo de modelos diagnósticos de evaluación, específicamente aquellos basados en aspectos cognitivos. De acuerdo con Gierl, Leighton y Hunka (2007) "una evaluación cognitiva diagnóstica 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).

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 (Coll & Onrubia, 2001).

Teoría del espacio del problema

El estudio de la resolución de problemas se describe en la obra magna de Newell y Simon (1972) titulada "Human Problem Solving", en la cual la *teoría del espacio del problema* continúa siendo el centro de la investigación actual en esa materia (Eysenck & Keane, 2005).

Básicamente, esta teoría conceptualiza la resolución de problemas como un proceso de búsqueda a través del *espacio del problema* mediante una ruta que conecta el *estado inicial de conocimiento,* entendido como la comprensión que tiene la persona de la información dada, con el *estado* final o solución deseada (Novick & Bassok, 2005).

El *espacio del problema* se refiere a la representación que la persona genera de la tarea que se le presenta. Dicha representación es decisiva en el proceso de solución, pues determina qué conocimiento se activará en la memoria a largo plazo y establece la complejidad cognitiva del problema, que está en función de la medida en la que el espacio refleja las características del *ambiente de la tarea.*

El concepto de *ambiente de la tarea* hace referencia a la descripción objetiva de la situación, esto es, el problema real tal y como sería descrito por alguna persona –observadora omnisciente– que dispone de todo el conocimiento relevante y de las condiciones óptimas de procesamiento para su solución (Simon, 1978).

Cuando se especifica el ambiente de la tarea de una actividad, deben identificarse: a) una descripción del estado inicial; b) una descripción del estado final o meta en el cual el problema estará resuelto; c) una serie de operadores o acciones posibles que sirven para alterar el estado actual del problema; d) un conjunto de limitaciones o condiciones adicionales acerca de cómo se deberá resolver el problema (Holyoak, 1995); e) la información que está al alcance de quien lo solucionará; f) las entidades que constituyen o hacen parte de la situación; y g) el conocimiento relevante (Newell & Simon, 1972).

La resolución de problemas desde la teoría del procesamiento de la información requiere de un sistema de procesamiento que genera estrategias en función de las demandas de la situación problemática (Simon, 1978) y que, por su carácter adaptativo, posibilita una correspondencia entre el espacio del problema y el ambiente de la tarea. La persona que solucionará el problema, al ejecutar los operadores o las acciones legales en una secuencia correcta, logrará resolverlo, para lo cual debe atravesar por una serie de nodos o estados intermedios.

De esta manera, el espacio de la tarea está constituido por *estados de conocimiento* (el estado inicial, el estado final y varios estados intermedios posibles), un conjunto de operadores que permiten el movimiento de un estado de conocimiento a otro y de información local acerca de la ruta que se ha tomado en el espacio, es decir, el estado actual de conocimiento y cómo se llegó a él. El estado actual de conocimiento es aquel en donde la persona que solucionará el problema se ubica en cualquier punto del proceso de solución (Novick & Bassok, 2005).

El análisis del ambiente de la tarea permite valorar racionalmente el nivel de dificultad de una situación o su complejidad cognitiva, esto es, sin necesidad de recurrir al desempeño real de las personas en la actividad. Sin embargo, esto solo da una visión parcial de las características de la ejecución. Es necesario llevar a cabo un análisis del espacio del problema; es decir, de la *representación* que la persona tiene de la tarea misma y que determina, según la teoría, la manera como efectivamente las personas transforman los estados del problema por medio de sus acciones.

Según Newell y Simon (1972), el proceso de solución de problemas incluye dos subprocesos: comprensión y búsqueda. Durante la *comprensión*, la persona identifica, por vías de reconocimiento o razonamiento, la meta y demás componentes del problema. Durante la *búsqueda*, la persona busca y selecciona entre un conjunto variablemente amplio de acciones o estrategias (ambas consideradas reglas, sean de carácter heurístico o algorítmico) aquellas capaces de reducir la distancia entre el estado inicial y el estado final. Tal conjunto de operaciones o acciones se conoce como *espacio de reglas*, las cuales pueden ser aplicadas en cada paso del proceso de búsqueda (Simon, 1978).

Como al aumentar el número de reglas o la profundidad de la búsqueda es posible exceder la capacidad de procesamiento impuesta por los límites de la memoria a corto plazo y la memoria a largo plazo, la solución inteligente de problemas considera solo unas pocas alternativas que resultan prometedoras. Por ello, la efectividad en la solución de problemas está estrechamente relacionada con el empleo de métodos que permitan limitar la búsqueda por el espacio del problema (Holyoak, 1995).

Investigaciones recientes sobre todo en el dominio de pensamiento científico (Li & Klahr, 2006; Klahr, 2000; Klahr & Simon, 1999) plantean que la búsqueda de la solución de un problema requiere de tres procesos interdependientes: el *espacio de hipótesis*, el *espacio de experimentos* y la *valoración de evidencias*. En el primero de ellos, el estado inicial consiste en

algún conocimiento que se tiene acerca del dominio específico y el estado final lo constituye una hipótesis que puede abarcar parte de dicho conocimiento o su totalidad. Cuando se está considerando una o más hipótesis, no es inmediatamente obvio que estas constituyan un experimento adecuado e informativo para la resolución de la situación particular. Finalmente, el tercer proceso involucra la comparación de las predicciones derivadas de la hipótesis en estudio con los resultados obtenidos a partir de la experimentación.

Dado que el enfoque del procesamiento de la información de Newell y Simon (1972) cambió significativamente la investigación sobre resolución de problemas, apartando la atención de las condiciones bajo las cuales se dan las soluciones y dirigiéndola hacia los procesos cognitivos y componentes implicados que transforman el estado inicial del problema para conseguir el estado final propuesto, es relevante estudiar la influencia que esta teoría ha tenido en dominios específicos, tales como la Matemática. La siguiente sección se dedicará a ello.

Capacidad matemática

Desde un punto de vista elemental, la Matemática se ha concebido como un conjunto de reglas y procedimientos para realizar cálculos (Resnick & Ford, 1990). Autores como Orton (1998) afirman que "las matemáticas son tanto un producto como un proceso; tanto un cuerpo organizado de conocimientos como una actividad creativa en la que participa el que aprende" (p. 51). De esta manera, la combinación de elementos del conocimiento, reglas, técnicas, destrezas y conceptos previamente adquiridos permite, a quien aprende, dar solución a una situación nueva, no rutinaria, es decir, resolver un problema.

Para lograr una adecuada competencia en la resolución de problemas, las personas tienen que adquirir habilidades procedimentales y comprender los conceptos y los principios que se relacionan con dichas habilidades (Bruning, Schraw & Ronning, 2009).

De acuerdo con Onrubia, Rochera y Barberà (2001), la Matemática se ha constituido en uno de los dominios 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.

Competencias implicadas en la capacidad matemática

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*.

Autores como Onrubia, Rochera y Barberà (2001) y Schoenfeld (2007) concuerdan en que hay cuatro capacidades básicas de la capacidad matemática: 1) el dominio de una amplia base de conocimiento declarativo o conocimiento base, 2) el dominio de un conjunto igualmente amplio de procedimientos específicos, así como la posibilidad de emplearlos estratégicamente, 3) la facultad de controlar metacognitivamente ambos tipos de conocimiento –usar con efectividad lo que se sabe–, y 4) poseer una determinada actitud, inclinación y sensibilidad hacia la Matemática.

La primera capacidad hace referencia al *conocimiento declarativo*, el cual incluye el conocimiento de hechos, conceptos y sistemas conceptuales y principios de carácter

matemático. Por ello, se encuentra fuertemente mediado por el tipo de lenguaje formal y por los sistemas notacionales en que se expresa.

De acuerdo con Onrubia, Rochera y Barberà (2001), el conocimiento declarativo

> aporta elementos relevantes que es preciso reconocer para ejecutar un procedimiento particular, como las características de un problema y sus condiciones internas. Así entendido, el conocimiento declarativo influye decisivamente en la comprensión y representación adecuadas y pertinentes de los problemas susceptibles de ser resueltos por métodos matemáticos, así como en la formación de nociones que posteriormente se utilizarán de manera aplicada. De no construirse de un modo sólido y congruente, estas nociones llevarán a los alumnos a errores importantes (p. 491).

En segundo lugar, el *conocimiento procedimental* supone la aplicación de secuencias de acciones y operaciones de las que se obtiene un resultado acorde con un objetivo particular.

Hay dos grandes tipos de procedimientos: los *algorítmicos* y los *heurísticos*. En el primer caso, conducen a una solución adecuada si se siguen todos los pasos prescritos; en el segundo caso, no garantizan una solución correcta, pero guían de manera sistemática el proceso para llegar a ella (Onrubia, Rochera & Barberà, 2001). Las estrategias ayudan a interpretar los problemas, a localizar el conocimiento y los procedimientos almacenados, por lo que organizan el proceso de pensamiento y permiten preparar un plan de acción que sea capaz de resolver la tarea planteada (Resnick & Ford, 1990). Consecuentemente, la interacción entre el conocimiento conceptual o declarativo y el conocimiento procedimental produce un dominio matemático cada vez más complejo (Bruning, Schraw & Ronning, 2009).

Tanto el conocimiento declarativo como el procedimental deben aplicarse intencional y conscientemente según las condiciones en que se desarrolla la acción. Esto posibilita que las personas que aprenden un procedimiento matemático particular, no lo apliquen de manera indiscriminada, es decir, se dispone de un sistema de valoración sobre la extensión y las limitaciones de lo que se sabe, para lo que se requiere de las *habilidades metacognitivas* de planificación, revisión, control, selección y evaluación de las propias actividades intelectuales (Onrubia, Rochera & Barberà, 2001; Schoenfeld, 2007).

Por último, hay una serie de aspectos afectivos, relacionales y motivacionales que son esenciales para el desarrollo de una capacidad matemática, tales como, actitudes positivas hacia la Matemática, desarrollo de buenos hábitos de trabajo y valoración de los fines, la potencialidad y la relevancia de este ámbito de conocimiento (Rico, Marín & Romero, 2001).

Una vez discutidos algunos aspectos propios de la capacidad matemática, se presentarán componentes de su estudio desde un enfoque de procesamiento de la información, dado que este trabajo de investigación particulariza en la resolución de problemas enfocada en la resolución de ítems de una prueba, así como por ser este el posicionamiento teórico adoptado por razones expuestas anteriormente.

De acuerdo con Mayer (1986) el estudio de la capacidad matemática de las personas se puede hacer desde dos aproximaciones básicas: el enfoque psicométrico y el enfoque del procesamiento de la información. El primero de ellos define este concepto como lo que mide una prueba de Matemática, esto es, la capacidad para resolver bien los ítems que conforman una prueba. No obstante, tal concepción proporciona una descripción que depende de lo que se está midiendo. Por su parte, el enfoque del procesamiento de la información concibe la capacidad matemática como el conjunto de operaciones cognitivas, habilidades y conocimientos que son componentes de las tareas matemáticas, para lo cual se basa en el

análisis de tareas. Este enfoque supera algunas de las dificultades que presenta la perspectiva psicométrica.

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. Este es el propósito del surgimiento de enfoques de evaluación con un claro énfasis en procesos cognitivos.

El método *rule space*

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.

Instituto de Investigación en Educación
Unidad de Información y Referencia

Una de esas propuestas se desarrolló en un programa de investigación dirigido por Kikumi Tatsuoka (1983, 1985, 1990, 1995, 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 (Rupp, 2007).

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).

El método *rule space* se considera útil en problemas de reconocimiento y clasificación estadística de patrones de respuesta y se le considera un enfoque determinista y probabilístico. En el primer caso, puesto que se hace referencia a la selección de variables constituyentes del modelo –los atributos– y en el segundo, dado que se lleva a cabo una clasificación de patrones de respuesta basada en las probabilidades de dominio de tales atributos.

Según Tatsuoka (2009),

> el método *rule space* transforma un conjunto de datos de estudiantes compuesto por puntuaciones de ítems en un conjunto de datos compuesto por probabilidades de dominio de atributos. El beneficio de este enfoque es que permite el diagnóstico del rendimiento de tales estudiantes en una prueba en términos detallados de habilidades de procesamiento y de conocimiento del contenido (p. 9).

De acuerdo con Rupp (2007), en la metodología *rule space*, el modelo de TRI que se emplee permite transformar las puntuaciones observadas en una escala latente continua con

propiedades matemáticas deseables que posibilitan comparaciones cuantitativas referidas a normas entre personas. A esto agrega que

> Numéricamente, se puede observar que los rendimientos globales o las características que la evaluación pretende medir son capturados, mediante una métrica amplia, mediante la creación de una variable latente continua en el modelo de TRI y mediante una métrica de granularidad más fina que es la creación de vectores discretos de atributos que contienen las especificaciones de atributos para los ítems (p. 17).

Tatsuoka (1987) mostró que los procesos cognitivos subyacentes a la resolución de un ítem son determinantes de la pendiente y de los parámetros de ubicación de una curva en un modelo de TRI. Por ello, en su método incluyó la utilización del "*índice de precaución*", denotado por ζ, para representar el patrón de respuestas de cada persona –no solo con el valor de θ– y con ello brindar información diagnóstica de mayor significado.

Si un patrón observado de respuestas tiene ceros para los ítems más difíciles y unos para los ítems más fáciles, entonces ζ asume un valor negativo; en caso contrario, ζ adoptará valores positivos cada vez mayores, con lo que se obtiene una medida de lo inusual de un patrón de respuestas. Sin embargo, se espera que el valor de ζ sea cero (Tatsuoka, 1996, 2009).

Representación de patrones de respuesta

Al aplicar consistentemente reglas erróneas o concepciones equivocadas, se determinan patrones de respuesta conocidos como "*patrones ideales de respuesta*" (Tatsuoka, 1983, 1996). El método *rule space*, por medio de funciones especializadas conocidas como

funciones booleanas descriptivas, logra describir el comportamiento ideal en términos de los atributos identificados previamente. Por ello, representa todos los patrones de respuesta (tanto los de las personas como todos los patrones ideales posibles) en un espacio de pares ordenados en el que es posible examinar la proximidad del punto en el que se ha representado una persona con respecto a los patrones ideales de respuesta, llamados *estados de conocimiento.*

La representación gráfica de los patrones ideales de respuestas a los ítems de una prueba se hace en términos de dos variables: θ (theta) y ζ (zeta). La primera de ellas corresponde al nivel de aptitud o habilidad según algún modelo de TRI y se utiliza para describir los patrones ideales de respuesta. De esta manera, una persona de alta habilidad tendría un patrón ideal 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 ideal de respuestas con la mayoría de elementos iguales a cero. Por otra parte, tanto una persona de habilidad alta que no acierte algunos ítems fáciles, como una de habilidad baja que acierte algunos ítems difíciles, tendrán una medida alta en la escala de respuestas inusuales o atípicos. Tal medida es ζ y constituye la segunda dimensión que el método *rule space* utiliza para describir las respuestas de las personas (Birenbaum, Kelly & Tatsuoka, 1992).

La representación gráfica de las variables continuas θ y ζ tiene interpretaciones con gran significado. En particular, los puntos ubicados en el extremo izquierdo del espacio tienen valores negativos de ζ, lo cual quiere decir que los patrones de respuesta a los ítems correspondientes a èstos 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, es decir, por lo general unos para los ítems más difíciles y ceros para los más fáciles (Tatsuoka, 2009).

El índice ζ ha sido utilizado en diagnósticos cognitivos sobre errores en que incurren las personas al responder los ítems de una prueba. Por ejemplo, Birenbaum y Tatsuoka (1983) investigaron la relación entre los patrones inconsistentes de respuestas y los errores en la aplicación de procesos cognitivos y determinaron que la aplicación consistente de reglas erróneas causa patrones inusuales de respuestas. De esta manera, la variable ζ es una medida de lo atípico que resultan ciertos patrones de respuesta.

De acuerdo con el método, lo deseable es que la cantidad de estados de conocimiento no sea muy grande, para que los resultados de la clasificación sean precisos. Por ejemplo, si dos estados de conocimiento no están bien separados, entonces los errores que se cometan al clasificar serían mayores. Una forma de evitar problemas de precisión en los resultados es separar a las personas en grupos de clasificación, lo cual es posible agregando ítems a la prueba o aumentando la dimensión del *rule space* (Tatsuoka, 1996).

Estados de conocimiento

Los estados de conocimiento –también llamados *estados latentes de conocimiento*– consisten en patrones binarios de atributos que expresan el dominio o no de tales atributos, por lo que están formados por unos y ceros. En consecuencia, si una persona puede utilizar correctamente un atributo k, entonces el k-ésimo elemento del estado de conocimiento es un 1, o de lo contrario, será un 0. Para ello se asume que la respuesta correcta para el ítem j se obtiene si y solo si todos los atributos involucrados en dicho ítem se han aplicado exitosamente. De este modo, la probabilidad de responder correctamente el ítem j se determina multiplicando las probabilidades del uso correcto de los atributos involucrados en el ítem (Tatsuoka, 2009).

Procedimiento metodológico

En este apartado se expone la forma en que se desarrolló el proyecto de investigación.

Tipo de investigación

En términos generales, el estudio se considera no experimental, dado que no se tiene control directo sobre las variables independientes debido a que sus manifestaciones ya han ocurrido, de ahí que solo sea posible hacer inferencias sobre las relaciones entre las variables, sin intervenir directamente en ellas. A estudios de este tipo se les conoce como investigaciones *ex post facto* (Kerlinger & Lee, 2002) y se les considera una herramienta muy valiosa para la explicación de situaciones, sobre todo por la utilización de técnicas estadísticas avanzadas (Cohen, Manion & Morrison, 2008).

En cuanto a la clasificación de la investigación según su objeto de estudio, se plantea que sea de tipo explicativo, pues se pretende proporcionar un sentido de entendimiento del fenómeno que se investiga por medio del establecimiento de las causas de los eventos que se estudian (Johnson & Christensen, 2008).

Consecuentemente, se propone que la presente investigación permita explicar los procesos, las estrategias y las estructuras de conocimiento de estudiantes costarricenses de undécimo año en la resolución de ítems de la prueba de bachillerato en Matemática correspondiente al año 2006 para colegios académicos diurnos. Lo anterior, se plantea desde una perspectiva psicométrico-cognitiva, específicamente con la utilización del método *rule space*.

Método empleado

Dado que el presente estudio se basa en una evaluación diagnóstica específicamente en el desempeño en una prueba de Matemática, se requiere un modelo cognitivo que permita explicar los datos. Por ello, la elaboración de una lista de atributos cognitivos, su validación y la matriz de incidencia fueron las primeras tareas que se desarrollaron. Posteriormente, para validar el modelo cognitivo, se llevaron a cabo los análisis propios del método *rule space* para obtener la información diagnóstica del desempeño del estudiantado en la prueba de bachillerato en Matemática mediante los vectores de probabilidad de dominio de atributos.

Como fase previa al seguimiento de las preguntas de investigación y para una mejor comprensión de los datos obtenidos en la aplicación de la prueba de bachillerato en Matemática que se llevó a cabo en noviembre 2006, se calcularon medidas tales como cantidad de personas que contestaron la prueba, valores máximo y mínimo de la distribución, promedio general en la prueba (la escala utilizada es de 0 a 100 puntos) y desviación estándar. Asimismo, se calculó el coeficiente de confiabilidad alfa de Cronbach y se determinó si existían ítems de los cuales fuera posible prescindir para maximizar tal coeficiente.

También se realizaron cálculos psicométricos basados en la teoría clásica de los test, el modelo de Rasch y el modelo de dos parámetros de la TRI. Para tales estudios se utilizó el software Winsteps 3.74.0 (Linacre, 2012b) y el BILOG-MG 3.0 (Zimowski, Muraki, Mislevy & Bock, 2003).

Descripción de las técnicas utilizadas

Con el propósito de llevar a cabo un análisis del ambiente de la tarea, se desarrollaron protocolos escritos por parte de dos especialistas. Según Tatsuoka (1990), se debe consultar a personas expertas y docentes con experiencia, de tal modo que si utilizan métodos diferentes para resolver un conjunto de problemas, entonces podrían obtener atributos distintos para los ítems.

Los protocolos de especialistas se desarrollaron con el propósito de documentar las representaciones y los procesos que tales personas consideran que son los que el estudiantado utiliza con mayor frecuencia para resolver problemas de las áreas temáticas medidas en la prueba. No obstante, dado que el análisis del ambiente de la tarea solo da una visión parcial de las características de la actividad, se hace necesario llevar a cabo un análisis del espacio del problema, lo cual se detalla en el siguiente apartado.

Con el fin de obtener evidencias acerca de las representaciones y los procesos utilizados por el estudiantado para resolver los ítems de la prueba, se utilizó la técnica de "pensar en voz alta" (Ericsson & Simon, 1993).

Con la aplicación de esta técnica se pretendió conseguir un análisis más profundo de los procesos cognitivos utilizados por una persona al responder un ítem y así lograr que las inferencias que se hagan del procesamiento cognitivo del estudiantado sean más sólidas (Desimone & Le Floch, 2004; Leighton, 2004).

La utilización de protocolos verbales es una de las formas de validación de las descripciones de los atributos de los ítems que conforman la prueba, con personas que sean comparables con la población que se investiga (Gierl, Leighton & Hunka, 2000). Estos protocolos, junto a los desarrollados por las personas especialistas, permitieron elaborar una

lista preliminar de atributos cognitivos, práctica que se evidencia en la literatura técnica del método *rule space* (Katz, Martinez, Sheehan & Tatsuoka, 1993; por ejemplo).

Una vez contactado el centro educativo y concluida la negociación de entrada con el director de la institución, el proceso de selección de las personas se delegó a la profesora de Matemática del nivel de undécimo año, a quien se le entregó un documento con la explicación de la técnica y su propósito. Asimismo, se le solicitó lograr una distribución equitativa según sexo. Se contó con la participación de seis estudiantes.

Previo al desarrollo de la técnica, el investigador se reunió con las personas seleccionadas en presencia del director para explicar en qué consistía el estudio y la trascendencia de su participación. En ese momento se les dio el formulario de consentimiento informado y se les solicitó que lo llevaran firmado el día que les correspondiera su turno, el cual fue escogido entre el director y la profesora.

El audio de las sesiones fue transcrito por el investigador y fue utilizado como insumo para la elaboración de la lista de atributos, así como para el manual para la codificación, lo cual se detallará más adelante.

Población a la que va dirigida la investigación

Como este proyecto de investigación se basa en el análisis de la estructura cognitiva de una prueba de gran escala, los datos provienen de la Prueba de Bachillerato en Matemática aplicada en noviembre del 2006. La cantidad de personas que presentaron el examen fue de 21488, provenientes de 704 centros educativos. Esta es una prueba de carácter censal y de altas consecuencias para la población estudiantil, en particular,

Instituto de Investigación en Educación
Unidad de Información y Referencia

constituye el 60% del cálculo de la nota que define su condición final para egresar de la educación secundaria y, eventualmente, continuar con su formación universitaria.

Por las características del presente estudio, se cuenta con la base de datos constituida por las respuestas dadas por las personas a los ítems de la prueba en forma dicotómica (1 en caso de acertar el ítem y 0 en caso contrario).

De esta manera, la investigación va dirigida tanto al estudiantado que concluye su educación preuniversitaria o que está por concluirla, así como a docentes de secundaria y de cursos básicos propios de la educación superior, pues contarán con una caracterización del conocimiento y las habilidades en Matemática adquiridas por el estudiantado. De igual modo, se espera que diversas personas investigadoras desarrollen y validen este modelo y propongan otros. Finalmente, va dirigida a distintas autoridades ministeriales y académicas, con el fin de que se tomen las medidas para que la interpretación de los resultados en una prueba se pueda enfocar en las formas en que las personas procesan la información y cómo la utilizan en la resolución de problemas y, con ello, contribuir al mejoramiento de los procesos de enseñanza y aprendizaje.

Selección de participantes

Para seleccionar a tales especialistas y conseguir con ello los protocolos escritos, se consideraron las siguientes características: 1) experiencia laboral docente no menor a 10 años, 2) experiencia comprobada en el campo de la evaluación de la Matemática, y 3) experiencia comprobada en elaboración de pruebas de rendimiento a gran escala.

La selección de las personas para participar en la técnica de pensamiento en voz alta se enfocó al cumplimiento de las siguientes características: 1) estar cursando undécimo año,

2) ser estudiantes pertenecientes al Ciclo Diversificado en el sistema educativo costarricense, 3) aceptar participar voluntariamente en el estudio, 4) mostrar facilidad en la verbalización de las acciones emprendidas en la resolución de ítems en Matemática, y 5) firmar el formulario de consentimiento informado, junto con la debida autorización de padres, madres o personas encargadas legalmente.

Descripción de procedimientos para recolección y análisis de la información

Con el propósito de lograr el primer objetivo de este proyecto de investigación, se hacía necesario el diseño de una lista de atributos cognitivos para la prueba de bachillerato en Matemática del año 2006 para colegios académicos diurnos, la cual fue la base para la realización del modelo de diagnóstico.

Para su elaboración se recurrió a un detallado análisis del ambiente de la tarea mediante los protocolos escritos de dos especialistas que fueron seleccionados según las características ya mencionadas. El autor de este trabajo les brindó una charla a esas personas para explicarles tanto el propósito general de la investigación, como la necesidad de que resolvieran la prueba con un alto grado de detalle, con el fin de poder tener mejores evidencias para proponer atributos involucrados en la resolución de los ítems.

Posteriormente, el investigador proporcionó algunos ejemplos de protocolos escritos y de análisis de ítems que fueron discutidos con las personas especialistas. La revisión de la documentación propuesta, así como los comentarios realizados permitieron una mejor comprensión de cómo proponer atributos para conformar finalmente una lista preliminar para la prueba en estudio.

En una segunda jornada de trabajo, cada persona especialista resolvió los ítems de la prueba paso a paso y dejaron constancia escrita del proceso de resolución empleado. Esto

se desarrolló en el mismo lugar, de manera individual y siempre con la presencia del investigador. Para esta sesión de trabajo, cada persona contó con el conjunto total de ítems y la tabla de especificaciones con que la prueba fue elaborada.

Una vez que se finalizó la resolución de todos los ítems de la prueba, se revisaron las respuestas entre las personas especialistas y el autor del presente estudio. Se realizaron comentarios acerca de los ítems y las formas de resolución, así como de la congruencia entre el ítem y la tabla de especificaciones. Posteriormente, se les solicitó escribir las habilidades, los procesos y los conocimientos involucrados en la resolución. Esto fue considerado como un primer insumo para elaborar la lista de atributos.

Al concluir la aplicación de la técnica de pensamiento en voz alta con el grupo de estudiantes que participaron se procedió a transcribir el audio de las sesiones de trabajo. Tanto con la transcripción como con las evidencias de la resolución de los ítems que cada estudiante dejó por escrito, se procedió a una segmentación de los protocolos y luego se realizó una codificación. Estas dos actividades son propias del análisis del espacio del problema y permitieron que las personas especialistas y el investigador pudieran contrastar el ambiente de la tarea con el espacio del problema y así conformar una lista preliminar de atributos.

A cada especialista se le entregó una descripción del conjunto inicial de atributos y se les solicitó que codificaran los ítems de acuerdo con esa lista, con el fin de tomar decisiones en cuanto a cuáles atributos representan las habilidades cognitivas que se requieren para responder correctamente los ítems de la prueba.

Cuando se completó la actividad descrita en el párrafo anterior, se realizó una plenaria entre las personas especialistas y el investigador con el fin de depurar las descripciones y con ello elaborar una lista de atributos para codificar los ítems de la prueba

y que permitiera conformar la matriz de incidencia necesaria para los análisis con el método *rule space*.

En este punto se requirió la participación de otras dos personas especialistas con las mismas características ya detalladas para quienes participaron en la elaboración del análisis del ambiente de la tarea. A estas dos personas se les entregó un manual de codificación en el que se indicaron los atributos, su definición y algunos ejemplos.

Estas personas resolvieron individualmente la prueba completa, el investigador revisó con cada una las respuestas y, posteriormente, se procedió a la codificación de los ítems con la ayuda del manual.

Con esta información, se procedió a determinar la confiabilidad de las codificaciones calculando el grado de concordancia entre evaluadores, esto es, el coeficiente Kappa de Cohen (κ) y el coeficiente Phi (ϕ); para depurar la versión preliminar del trabajo. En caso de existir discrepancias, fueron sometidas a discusión y se promovió el consenso como forma de llegar a un acuerdo final en la codificación.

Para validar la lista de atributos y de la matriz Q se hicieron los siguientes análisis de regresión múltiple: 1) sobre los parámetros de dificultad con los vectores de atributos; 2) sobre las puntuaciones totales de las personas y las probabilidades de dominio de atributos; y 3) sobre el nivel de aptitud de cada persona según el modelo escogido de TRI y las probabilidades de dominio de atributos. Los valores de R^2 obtenidos indican el porcentaje de la varianza del desempeño del estudiantado en la prueba que se explica por el dominio de atributos.

El método *rule space* transforma las puntuaciones de los ítems, que por definición son variables discretas, en las variables θ y ζ que son continuas y que se utilizan como coordenadas del espacio bidimensional de clasificación. Los análisis propios del método

Instituto de Investigación en Educación
Unidad de Información y Referencia

rule space se llevaron a cabo con el software PMAIN-RULE SPACE (Tatsuoka, Varadi & Tatsuoka, 2004). Particularmente, se dispuso de una entrada de datos (la matriz *Q*), los patrones de respuesta de las personas a los ítems y las estimaciones de los parámetros de dificultad y discriminación propios de los modelos de TRI. Con la lista de atributos definitiva o final se determinaron las habilidades cognitivas, los procedimientos y los conocimientos subyacentes a la resolución de los ítems de la prueba.

Para lograr el segundo objetivo, en el que se pretende establecer los estados de conocimiento más comunes que describen los patrones de estructuras de conocimiento o atributos cognitivos de las personas que contestaron la prueba de bachillerato en Matemática 2006, se recurrió a diversas técnicas estadísticas.

Dado que un estado de conocimiento describe el patrón de dominio y no dominio de los atributos considerados como relevantes, entonces es posible obtener 2^n posibles estados de conocimiento si se han incluido n atributos.

Como el método *rule space* utiliza una regla de decisión bayesiana para encontrar los estados de conocimiento más próximos al punto en el que se representan las respuestas de una persona en un espacio de reglas, se debe tomar la decisión acerca de qué tan cerca está cada patrón de respuestas de la persona respecto de un estado predeterminado de conocimientos.

Para lograr tal cometido, se calculó la distancia de Mahalanobis (D^2) entre el patrón observado de respuestas de cada persona y un patrón ideal o esperado. El punto de corte para D^2 se estableció en 4,5 (Chen, 2006; Dogan, 2006; Tatsuoka, 2009).

Siguiendo recomendaciones técnicas (Tatsuoka, 2009) y estudios previos (Chen, 2006; Dogan, 2006) se consideró satisfactoria una tasa de clasificación mayor o igual que el

90% de la cantidad de personas involucradas en el estudio. Cabe destacar que esta tasa de clasificación también está vinculada con la validez de la lista de atributos y con la matriz Q.

Con la ayuda del software PMAIN-RULE SPACE (Tatsuoka, Varadi & Tatsuoka, 2004), se obtuvieron los diez estados de conocimiento más próximos a cada persona. Posteriormente se determinó cuáles estados de conocimiento tenían una frecuencia mayor o igual a 1% como tasa de clasificación de personas para ser considerados como estados comunes de conocimiento (Dogan, 2006).

De esta manera, se tuvo un procedimiento final de clasificación de personas, por lo que se determinó el número de estados de conocimiento y el porcentaje de estudiantes clasificados.

Análisis y discusión de los resultados

En este capítulo se organiza la información recopilada y se discuten los resultados propios de la investigación.

Características de los ítems y de la prueba

Los datos provienen de la Prueba de Bachillerato en Matemática aplicada en noviembre del 2006. En este caso se tiene N = 21 461 registros, cada uno con 60 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 los ítems. Por la cantidad de ítems, la escala va de 0 a 60 puntos. La puntuación total sigue aproximadamente una distribución normal con una media aritmética de 34,77 y una desviación estándar de 8,61.

La revisión de aspectos técnicos de la prueba posterior a su aplicación resultó en la eliminación de 3 ítems: uno de ellos (el 15) por presentar problemas de elaboración de la figura, otro (el 24) por no tener respuesta correcta en las opciones y el tercero (el 52) por presentar dos respuestas correctas. No obstante, dado que los resultados de esta prueba tienen altas consecuencias para las personas, en la práctica la puntuación total se calcula asignando un valor de 1 para tales ítems, puesto que una disminución en la escala resultaría perjudicial para el cálculo de la nota final.

Por otra parte, se determinó la fiabilidad de los resultados para los 57 ítems con la ayuda del software SPPS Statistics 16.0 (SPSS, 2007) para el cálculo del alfa de Cronbach y se encontró $\alpha = 0{,}860$ lo que indica que hay alta consistencia interna.

Instituto de Investigación en Educación
Unidad de Información y Referencia
INIE

Con el SPSS 16.0 se realizaron análisis para reducir la escala, esto es, la posibilidad de eliminar algún ítem si con eso se contribuye a una mejor consistencia interna en los resultados de la prueba.

Como resultado del análisis anterior, el alfa aumenta ligeramente su valor si se eliminan los ítems 41 y 59; asimismo, dichos ítems muestran correlaciones casi nulas con el total de la prueba. En este momento del estudio, se toma la decisión de mantener ambos ítems para llevar a cabo los análisis psicométricos por dos razones: 1) el aumento del alfa no se considera relevante, y 2) para no perder más elementos que conforman la prueba, sobre todo porque ya se tuvieron que eliminar tres ítems por fallas técnicas y porque estos dos pertenecen a temáticas relevantes objeto de medición. Sin embargo, mantenerlos en los análisis propios del método *rule space* dependerá de su ajuste al modelo de TRI escogido.

A pesar del alto valor del alfa, se evaluó la dimensionalidad de la prueba mediante un análisis factorial con el método de componentes principales y con rotación promax.

La medida de adecuación muestral de Kaiser-Meyer-Olkin (KMO) es 0,943 y el índice de prueba de esfericidad de Bartlett $\chi^2(1596, N = 57) = 145377{,}48; p = 0{,}000$ los cuales permiten concluir que el análisis factorial es adecuado para reducir la dimensionalidad de la escala.

El porcentaje de varianza explicada en el primer componente o factor es de 12,46% y en el segundo componente o factor es de 5,45%. Es decir, el 17,91% de la variabilidad observada en los ítems se explica con los factores comunes y el resto es variabilidad propia de cada ítem en particular. Para estos dos factores, los autovalores son 7,10 y 3,11 respectivamente.

De acuerdo con la matriz de correlación de factores, hay evidencia de que los componentes muestran una asociación relevante (0,352) por lo que se acepta un análisis conjunto de todos los ítems.

En cuanto a las características de los ítems de la prueba, se calcularon los índices de dificultad y de discriminación según la TCT y según modelos de TRI: el modelo de Rasch y el modelo de dos parámetros. En particular, los cálculos de TCT se llevaron a cabo con el SPSS; los del modelo de Rasch, con el software Winsteps 3.74.0 (Linacre, 2012b) y los correspondientes al modelo de dos parámetros, con el software BILOG-MG 3.0 (Zimowski, Muraki, Mislevy & Bock, 2003).

En el caso de la teoría clásica, la proporción de respuestas correctas varió de 0,09 a 0,93 con una media aritmética igual a 0,60 y una desviación estándar de 0,21. En cuanto a la discriminación, los valores oscilaron entre 0,12 y 0,42 con un promedio de 0,30 y una desviación estándar de 0,08.

Según el modelo de Rasch, el valor de θ varió de –2,49 a 3,41, con un promedio de 0 y una desviación estándar de 1,26. Finalmente, con el modelo logístico de dos parámetros el valor de θ varió de –3,40 a 3,83.

Para los siguientes análisis se tomó la decisión metodológica de trabajar con una muestra, sobre todo porque los programas de cómputo disponibles actualmente para este tipo de análisis tienen algunas limitaciones en la cantidad de cálculos que ejecutan. Por ejemplo, cuando se tiene una gran cantidad de atributos (18 en el caso de esta disertación) se tendrán 2^{18} probabilidades de combinaciones y los cálculos estarán basados en la creación de una matriz cuyo orden está determinado por la cantidad de personas y dichas probabilidades.

De esta manera, los cálculos se realizarán a partir de una matriz de orden 384×2^{18}, lo cual es un número suficientemente grande para ejecutar los algoritmos programados (A. Robitzsch, comunicación personal, 18 de septiembre, 2013).

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.

Nuevamente se contabilizaron el mínimo, el máximo, la media aritmética y la desviación estándar de la proporción de respuestas correctas y de θ según el modelo de Rasch y el modelo logístico de dos parámetros de la TRI.

En este caso, para la TCT se obtuvo que la proporción de respuestas correctas varió de 0,112 a 0,934, con un promedio de 0,597 y una desviación estándar de 0,207. En cuanto al modelo de Rasch, el valor de θ varió de –2,41 a 3,19, con un promedio de 0 y una desviación estándar de 1,20. Finalmente, con el modelo logístico de dos parámetros el valor de θ varió de –3,47 a 3,66.

Es de notar la similitud entre los estadísticos para los tres modelos en la población y en la muestra, por lo que la muestra se considera representativa de la población y se harán los cálculos basándose en este subconjunto.

Dada la variedad de la información que es posible obtener con el modelo de Rasch, se eligió para hacer los análisis propios del *rule space*, aunque una vez conformada la matriz Q, se hicieron análisis de regresión con los tres modelos, con el fin de tener mayores evidencias de la varianza explicada por los atributos propuestos.

Distribución de las puntuaciones obtenidas

El 25% del estudiantado contestó correctamente 27 ítems o menos, el 50% acertó 33 ítems o menos y, finalmente, el 75% contestó correctamente 39 ítems o menos. La

puntuación mínima es 9 y la máxima es 53, por lo que no hay personas que hayan acertado la totalidad de los ítems de la prueba ni tampoco hay personas que los hayan fallado todos.

Análisis con el modelo de Rasch

Para llevar a cabo los análisis propios con el modelo de Rasch, se suprimieron los ítems con fallas técnicas (15, 24 y 52), los cuales se llevaron a cabo con el software Winsteps 3.74.0. Para la población se obtuvo una media aritmética de 33,2 y una desviación estándar de 8,4 (en una escala de 0 a 57 puntos).

Para efectos del modelo, el error estándar asociado a la estimación de la media aritmética es de 0,33. En cuanto al ajuste interno ponderado por la información (INFIT), se tiene una media cuadrática de ajuste interno (MNSQ) igual a 0,99 y en cuanto al ajuste externo (OUTFIT) con un valor de MNSQ igual a 1,03; ambos en la media aritmética de la escala global de la prueba.

La raíz del error cuadrático medio (REAL RMSE) calculado sobre todas las puntuaciones de las personas es de 0,34, con una desviación estándar de 0,84. Este error es un indicador del error total o varianza aleatoria y se calcula sobre la base de que los desajustes se deben a desviaciones de los datos hacia el modelo. Por su parte, (MODEL RMSE) asume que todo desajuste en los datos es un reflejo de la naturaleza estocástica del modelo; en este caso se obtuvo 0,33 con una desviación estándar de 0,84.

Una de las informaciones que ofrece el Winsteps es el ordenamiento de ítem según su medida. Para los ítems analizados, el que resultó de mayor dificultad es el 41 y el más fácil fue el ítem 1. La medida en lógitos [logits] de la dificultad de cada ítem va desde 3,41 lógitos a –2,49 lógitos. Se observan 33 ítems por encima de la dificultad media y 24 ítems por debajo de la dificultad media. El error estándar oscila entre 0,03 y 0,01.

Instituto de Investigación en Educación
Unidad de Información y Referencia
UNE

En cuanto a las medidas de ajuste interno (INFIT) y externo (OUFIT) cabe destacar su cercanía a la medida extrema superior que presentan los ítems 41 y 59. El ítem 41 muestra un INFIT MNSQ de 1,22 y un OUTFIT MNSQ de 2,20. El ítem 59 muestra un INFIT MNSQ de 1,28 y un OUTFIT MNSQ de 1,47. Dado que en la literatura se registra que números mayores o cercanos a 1,30 en tales valores indican demasiada aleatoriedad en los datos observados, se procede a suprimirlos del análisis. La correlación punto-medida (PTMEA CORR) es negativa para el ítem 41 y muy cercana a cero para el ítem 59. Esta correlación mide el grado de asociación entre la puntuación particular observada para el ítem y la puntuación total en la prueba y se ha utilizado en este caso como un criterio técnico que respalda la decisión de suprimir los dos ítems en cuestión para los análisis sustantivos de esta investigación.

En cuanto a la dimensionalidad de la prueba, se reporta que la varianza empírica explicada por las medidas es de 29,3% y la esperada por el modelo en caso de ajuste perfecto es de 29,1%, por lo que hay una ligera discrepancia entre el valor empírico y el esperado. En el caso de la varianza sin explicar es de 70,7% empíricamente y de 70,9% según lo esperado por el modelo de Rasch.

Linacre (2012a) indica que Winsteps descompone la varianza sin explicar utilizando el análisis de componentes principales, para lo que reporta a lo sumo cinco contrastes (interpretados como subdimensiones). Tales contrastes pueden deberse a un efecto aleatorio en los datos y, por ello, ser algo completamente accidental.

Winsteps ofrece una lista ordenada de los ítems según su dificultad de mayor a menor. El reporte generado identifica al ítem 41 como el más difícil y al ítem 1 como el más fácil. Asimismo, el software detecta a las personas con mayor desajuste al modelo de Rasch. En este caso, la persona con ID = 158 es la que presenta un mayor desajuste, puesto que acertó los ítems 39, 23, 32, 15, 44 y 21, los cuales se esperaba que fallara, dada su ubicación en la escala de habilidad. Por otra parte, de acuerdo con el modelo se esperaba que no

acertara los ítems 55, 6, 18 y 5, pero sí los acertó. Para esta persona, el valor del INFIT es 1,4 y el del OUTFIT es 3,6.

Otro caso relevante para ejemplificar es el de la persona con ID = 21. En este caso, según el modelo se esperaba que acertara los ítems 22 y 6; no obstante, esa persona los contestó incorrectamente. Sus valores de INFIT y OUTFIT son 0,9 y 1,9, respectivamente.

Elaboración de la matriz de incidencia

Para determinar las habilidades, los conocimientos y los procesos cognitivos medidos en la prueba, se utilizó la técnica de pensamiento en voz alta con estudiantes que cursaron undécimo año y protocolos escritos de especialistas en la enseñanza de la Matemática.

En cuanto a los protocolos de estudiantes, se les solicitó la resolución en voz alta y que quedara evidencia escrita del procedimiento empleado. Se procuró en todo momento que el investigador interviniera lo menos posible, salvo en casos en que hubiera silencios durante un largo periodo de tiempo o si se advertía que la persona que resolvía los ítems no encontraba "la salida"; en este caso, se consideró que lo más importante para el estudio era ofrecerle alguna sugerencia para encaminarle hacia la solución.

Una vez que se transcribió el audio de las sesiones por cada estudiante, se procedió a la segmentación del texto en componentes que evidenciaran las habilidades, los conocimientos y demás procesos involucrados en la resolución de cada ítem.

La transcripción de las sesiones de cada estudiante y el protocolo escrito de especialistas en enseñanza de la Matemática se utilizaron de forma complementaria para la elaboración de una lista preliminar de atributos. Se pretendió con esto complementar ambos análisis de la tarea (el análisis racional y el análisis del ambiente).

La primera lista elaborada estuvo compuesta por 21 atributos. En una sesión de trabajo con dos especialistas en enseñanza de la Matemática y elaboración de pruebas se procedió a elaborar la descripción de cada atributo, con el propósito de contar con una guía para la codificación de los ítems y con ello, la conformación de una matriz de incidencia en la que se reflejara la relación entre los ítems y los atributos pertenecientes a estos.

Con la lista inicial de atributos se procedió a la codificación de los ítems por parte de dos especialistas en Matemática, quienes analizaron cada ítem en función de los atributos involucrados en su resolución.

Primero se codificaron 10 ítems y se compararon las de ambas personas con el fin de asegurar la comprensión de lo que cada atributo implicaba. El autor de esta investigación también codificó los ítems con el fin de guiar la discusión y de enriquecer y afinar la definición de los atributos.

Una vez que se llegó a un consenso en el significado y el alcance de los atributos de la lista, se procedió a la codificación de todos los ítems de la prueba. Para ello, se solicitó a cada persona que escribiera a la par de cada ítem las letras "AT" seguida del número con que se identificó el atributo en la lista dada anteriormente

Con la guía elaborada para tal efecto, se procedió a la codificación de los 55 ítems y se procedió a realizar un primer análisis de regresión múltiple, considerando como variable dependiente el valor de θ (con el modelo de Rasch y el logístico de dos parámetros) y de proporción de respuestas correctas. En ninguno de los casos el porcentaje de varianza explicada en la dificultad del ítem llegó al 70% con la estructura de atributos propuesta.

Se llevaron a cabo nuevas revisiones de la lista preliminar de atributos, así como de guía para la codificación. Esto ocasionó otras propuestas por parte de las personas especialistas consultadas para tal efecto. Las listas siguientes variaron en tamaño (13 atributos, 10 atributos y 19 atributos). Para cada una de ellas se realizaron los análisis de

regresión múltiple sin llegar en ningún momento a obtener el mínimo recomendado en la literatura.

Matriz de incidencia y su confiabilidad

Una nueva revisión de los trabajos descritos anteriormente, permitió contar con una lista de 18 atributos. Este trabajo, desarrollado varios meses después de las sesiones anteriores, permitió una reconsideración de lo codificado e incluso una depuración de la guía de descripción de los atributos.

Con ello, los dos especialistas codificaron independientemente los 55 ítems de la prueba según la lista de atributos dada (con la descripción de cada atributo). La fiabilidad de la codificación se presenta en la tabla 3. Dicha fiabilidad se reporta con tres medidas: 1) el porcentaje de acuerdo, 2) el coeficiente Phi, el cual es un caso particular del coeficiente correlación producto-momento de Pearson para medir la asociación entre dos variables dicotómicas, y 3) el coeficiente Kappa de Cohen, que se puede interpretar como la proporción de clasificaciones consistentes más allá de las que se podrían esperar por azar.

Los resultados de dichos índices muestran un alto valor de fiabilidad como acuerdo entre jueces.

Instituto de Investigación en Educación
Unidad de Información y Referencia
UNE

Lista de atributos y su descripción

A continuación se expone la lista de los 18 atributos con su correspondiente descripción. Esta lista fue la guía empleada para la codificación de los ítems que componen la prueba.

AT1: Utilización de métodos de factorización

Empleo de métodos tales como inspección, fórmula general, diferencia de cuadrados, factor común y agrupación para determinar los factores de trinomios de segundo grado con una variable y de polinomios de tres y cuatro términos con una o dos variables. Se pueden aplicar uno o varios métodos para determinar tales factores.

AT2: Revisión de opciones

Utilización de las opciones del ítem para que, al revisarlas, verificarlas o evaluarlas, se pueda llegar a la respuesta. Esto sucede especialmente en los ítems en los que se tiene más de una expresión que compone la respuesta o se utilizan expresiones equivalentes pero que se presentan con algún tipo de manipulación algebraica, logarítmica o trigonométrica.

AT3: Simplificación de expresiones

Empleo de estrategias algebraicas, logarítmicas o trigonométricas para expresar este tipo de objetos matemáticos en su forma más reducida (forma canónica). Incluye efectuar operaciones de cualquier tipo con diversas expresiones para llegar a la forma más simple en el resultado.

AT4: Resolución de ecuaciones o desigualdades

Empleo de diversas estrategias para hallar la solución de una ecuación determinada o de un sistema de ecuaciones o desigualdades. Incluye el despeje de variables en una expresión matemática cualquiera.

AT5: Traducción y formulación de expresiones para resolver un problema

Establecimiento de expresiones que modelan matemáticamente una situación dada, esto es, la utilización del lenguaje simbólico para representar una situación y el establecimiento de la relación que vincula dos cantidades.

AT6: Interpretación de la información contenida en el enunciado

Lectura y comprensión del texto propio del enunciado del ítem el cual no es trivial, sino que contiene información relevante que debe ser organizada para identificar los datos existentes, la incógnita y los elementos contextuales de la situación, así como para decidir la estrategia de resolución que se empleará.

AT7: Determinación o cálculo de conceptos básicos de funciones

Identificación y cálculo de conceptos básicos de funciones (dominio, codominio, ámbito, imagen, preimagen) y de relaciones que se establecen entre conjuntos numéricos cuyo criterio está formulado mediante expresiones algebraicas. Incluye la representación gráfica y el análisis de esta forma de representación, así como criterios para clasificar funciones (según el criterio o el codominio).

AT8: Verificación de proposiciones

Empleo de diversas estrategias de razonamiento lógico inductivo o deductivo para determinar la veracidad de proposiciones dadas en el enunciado del ítem.

AT9: Utilización de la información explícita o implícita en una figura

Empleo de definiciones, axiomas y teoremas para determinar la información relevante de una figura, sin quedarse en la visualización de esta. Incluye el establecimiento de relaciones entre los componentes de la figura y eventualmente la necesidad de construir algún objeto matemático que no se encuentre dado explícitamente en la figura pero que se requiera para resolver el ítem; tal construcción dependerá del sistema axiomático y teórico propio del conocimiento requerido en el ítem.

AT10: Distinción entre variables y parámetros

Utilización del contexto propio del enunciado del ítem para discriminar entre variables (cantidades cuyo valor no es fijo) y parámetros (cantidades cuyo valor es fijo pero no conocido).

AT11: Utilización de fórmulas

Instituto de Investigación en Educación
Unidad de Información
Herencia

Empleo de fórmulas para resolver una situación dada, ya sea algebraica (cálculo del discriminante, por ejemplo), geométrica (áreas, perímetros y volúmenes) o relacionada con funciones (determinación del punto máximo o mínimo de una función cuadrática, cálculo de la pendiente de una o más rectas, por ejemplo).

AT12: Determinación o cálculo de conceptos de funciones exponenciales y logarítmicas

Identificación y cálculo del criterio, dominio, codominio, ámbito, cálculo de imágenes y preimágenes, variación, intersecciones con los ejes, representación gráfica. Incluye el establecimiento de características de estas funciones según la base propia del criterio.

AT13: Utilización de propiedades de potencias o logaritmos

Empleo de propiedades de potencias, de exponenciales y de expresiones logarítmicas en el cálculo, la simplificación y la resolución de situaciones que requieran de estos conceptos.

AT14: Determinación de elementos o propiedades de una figura plana

Identificación de objetos matemáticos propios de la geometría plana, tales como circunferencias concéntricas, circunferencias tangentes interiores y exteriores, circunferencias secantes, ángulos en la circunferencia, polígonos regulares: ángulos, lados, diagonales, radio, apotema.

AT15: Determinación de elementos o propiedades de dos o más figuras planas o un cuerpo geométrico

Identificación de elementos o propiedades de objetos propios de dos o más figuras planas en un misma figura o por aparte, o bien, de cuerpos geométricos, así como el establecimiento de relaciones entre tales elementos o propiedades. Incluye la determinación de la circunferencia inscrita o circunscrita a un polígono regular y sus relaciones con el polígono circunscrito o inscrito. Asimismo, el ítem puede incluir elementos o propiedades

de cuerpos geométricos: esfera, cubo, prisma y pirámides rectas, cono y cilindro circulares rectos.

AT16: Cálculo de áreas o volúmenes de figuras o regiones

Determinación del área del círculo, la longitud de la circunferencia o de un arco, así como la utilización de diversas estrategias para determinar áreas y perímetros del anillo o corona circular, del sector circular y del segmento circular. Incluye el cálculo del área de la base, el área lateral, el área total y el volumen de cuerpos geométricos.

AT17: Establecimiento de equivalencias entre expresiones trigonométricas

Utilización de relaciones trigonométricas recíprocas, fundamentales y para ángulos complementarios en el establecimiento de identidades trigonométricas.

AT18: Determinación o cálculo de conceptos de funciones trigonométricas

Identificación del criterio, el dominio, el ámbito de las funciones seno, coseno y tangente, así como determinación de imágenes y características propias de estas funciones: variación en el signo, representación gráfica, periodicidad, intervalos de monotonía, intersección con los ejes, puntos de discontinuidad (en el caso de la función tangente).

Con la referencia dada en la guía para la codificación y una vez que se recolectaron evidencias de la consistencia entre jueces, se procedió a elaborar la matriz Q.

Matriz de incidencia o modelo de diagnóstico

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

Clasificación previa según el método *rule space*

Una vez que se ejecutaron los análisis propios del método *rule space*, de las 384 personas que integran la muestra del estudio, se llega a una clasificación de 378, es decir, una tasa de clasificación del 98,4%, con un criterio de $D^2 < 4,5$ para la clasificación de 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 (Chen, 2006; Dogan, 2006).

Las seis personas que no pudieron ser clasificadas según el *rule space*, se identificaron con los números: 105, 153, 188, 198, 332 y 367.

Casi todos los patrones observados de respuesta corresponden a los estados de conocimiento predeterminados por el método *rule space* a partir de la matriz *Q* propuesta. Sin embargo, los patrones de estas personas no resultan completamente consistentes con ninguno de los estados de conocimiento generados y por ello sus respuestas no son comparables ni se aproximan a los patrones ideales o esperados de respuestas según el modelo. 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.

Estadísticas descriptivas de las probabilidades de dominio de atributos

Tomando como base los patrones de respuesta a los ítems de la prueba y el modelo cognitivo propuesto en la matriz *Q*, el método *rule space* brinda las probabilidades de dominio de atributos para cada estudiante que se pueda clasificar en alguno de los estados de conocimiento. Por ello, se trabajará con 378 personas que son las que resultaron clasificadas en al menos un estado predeterminado de conocimiento.

Un criterio muy frecuente en la literatura sobre *rule space* (Dogan, 2006; Chen, 2006) es la utilización de las *probabilidades de dominio de atributos* [attribute mastery probabilities] para validar la matriz *Q*. En este caso, se llevó a cabo un análisis de regresión múltiple con la puntuación total en la prueba como variable dependiente y las probabilidades de dominio de atributos como variables independientes. Se obtuvo un R^2 igual a 0,901 y un R^2 ajustado igual a 0,896; esto indica que el 90% de la varianza en las puntuaciones totales de la prueba se explican por las probabilidades de dominio de atributos, por lo que se considera un criterio adicional para establecer la validez de la matriz de incidencia propuesta para la prueba analizada.

Las probabilidades de dominio de atributos constituyen uno de los principales aportes del *rule space* en términos de la información diagnóstica que se pueda ofrecer individualmente a cada persona. El promedio de dichas probabilidades representa la información diagnóstica en términos generales, esto es, una visión global del conjunto de estudiantes que contestaron la prueba.

Los dos atributos con las más bajas probabilidades de dominio son el AT8 y el AT15; el primero de ellos se vincula al formato del ítem (puesto que se debe verificar la veracidad de proposiciones dadas en el enunciado) y el segundo es propio del área de la geometría. Asimismo, las probabilidades más altas se dan en los atributos AT3, AT2, AT4, AT1 y AT11 los cuales evidencian una fortaleza de la muestra en los atributos referidos a algoritmos.

Instituto de Investigación en Educación
Unidad de Información y Referencia
UNA

Una de las razones fundamentales de la utilidad del método *rule space* en las evaluaciones cognitivas diagnósticas es que brinda información individual y grupal acerca de las probabilidades de dominio de atributos, los cuales reflejan las estrategias, los procedimientos e incluso los conocimientos involucrados en la resolución de los ítems de una prueba en particular.

A diferencia de la TCT o de los modelos de TRI que indican de alguna manera lo que la persona hizo en la prueba, el *rule space* da información acerca de lo que esas personas pueden hacer y permite establecer mediaciones didácticas basadas en criterios empíricos con el fin de lograr el dominio de los atributos que la persona aún no ha logrado alcanzar; asimismo, da cuentas acerca de las habilidades ya dominadas.

Conclusiones y recomendaciones

Al investigar las habilidades cognitivas específicas, los procedimientos y los conocimientos que subyacen a la prueba de Bachillerato en Enseñanza Media del 2006 en la asignatura de Matemática, los resultados de la validación y los análisis de clasificación propios del *rule space* indican que los 18 atributos propuestos brindan evidencias para explicar las habilidades, los procesos cognitivos y conocimientos que se requieren para contestar los ítems de la prueba.

La cantidad de ítems vinculados a los atributos empleados en su resolución varía desde 3 hasta 18. Específicamente, la resolución de ecuaciones o desigualdades (AT4) es el atributo más utilizado en los ítems y el establecimiento de equivalencias entre expresiones trigonométricas (AT17) es el menos utilizado. Cabe mencionar que los atributos: revisión de opciones (AT2), simplificación de expresiones (AT3), interpretación de la información contenida en el enunciado (AT6) y utilización de fórmulas (AT11) aparecen entre 13 y 15 ítems cada uno; esto sugiere un marcado énfasis en la utilización de algoritmos para la resolución de los ítems de la prueba.

Por otra parte, solo el ítem 41 requirió de 6 atributos para su resolución, lo cual es coherente con la estimación del modelo de Rasch que permitió considerarlo como el ítem más difícil de la prueba. De igual manera, el ítem considerado más fácil únicamente requiere 2 atributos.

Dado que el método *rule space* se basa en modelos de TRI, se escogió el modelo de Rasch, puesto que brinda información de suma relevancia para propósitos diagnósticos (respuestas inesperadas, patrones que evidencian desajustes con respecto al modelo, por ejemplo).

Al considerar el nivel de habilidad (θ) según el modelo de Rasch como variable dependiente y los atributos cognitivos expresados en forma binaria como variables independientes se logró explicar más del 80% de la varianza en la dificultad de los ítems.

Instituto de Investigación en Educación
Unidad de Información

A pesar de que el modelo de Rasch se ha utilizado para evaluaciones diagnósticas, una de sus principales limitaciones es que todas las personas con la misma puntuación total tienen la misma puntuación en la escala de TRI, lo cual se conoce como *"suficiencia de la puntuación sumada"* (Martínez, 2005). Este aspecto no permite diferenciar cuáles son las fortalezas y las debilidades de las personas.

Los hallazgos confirman la limitación del modelo de Rasch para brindar información diagnóstica, puesto que ubica a estos dos grupos de personas cada uno en el mismo punto de la escala, a pesar de que son muy claras las diferencias observadas a partir de sus patrones de respuesta y del dominio de atributos. No obstante, el *rule space* ofrece un nivel de detalle individual y grupal que no solo reconoce estas disimilitudes, sino que también orienta la elaboración de estrategias didácticas para fortalecer las habilidades que las personas no dominan, incluso de una manera particular para cada una.

Como ejemplo de ello, se han mostrado trayectorias de aprendizaje que muestran redes de estados de conocimiento y en las que se visualiza claramente las posibles formas de lograr que el estudiantado domine los atributos según las probabilidades que el modelo establece. De igual manera, las rutas de mejoramiento permiten particularizar el diagnóstico, puesto que el establecimiento de las relaciones entre estados de conocimiento es de carácter individual. Con ello, el personal docente cuenta con bases empíricas para la elaboración de estrategias y actividades que propicien o fomenten el domino de los atributos por parte de cada estudiante.

Por otra parte, la tasa de clasificación para las personas que integran la muestra se considera satisfactoria, dado que solo 6 de las 384 no se pudieron clasificar en los estados predeterminados de conocimiento, es decir, se logró clasificar al 98,4% del estudiantado.

Como limitaciones del presente estudio, se tiene la utilización de una base de datos correspondiente a la prueba de bachillerato en Matemática para colegios académicos diurnos en el año 2006, por lo que el conjunto de atributos identificados y que constituyen el modelo cognitivo propuesto no necesariamente resulten apropiados para otras modalidades (académicos nocturnos, técnicos) del sistema educativo costarricense.

Una de las limitaciones más grandes para la realización de investigaciones como esta y que pueden explicar la carencia de un uso más difundido en la actualidad del *rule space* en el campo de la evaluación cognitiva diagnóstica es que solo existe un software para llevar a cabo los análisis, no se cuenta con un manual técnico o guía para su utilización y se ejecuta únicamente en Linux, con una interfaz que no es fácilmente comprensible. Asimismo, la licencia solo está disponible para propósitos de investigación académica y algunos algoritmos propios del método son prácticamente indescifrables, dado que no se han dado a conocer públicamente.

Para subsanar el faltante de software para trabajar con modelos de evaluación cognitiva se debe trabajar con varias aplicaciones (SPSS, Excel, SAS, Mplus, por ejemplo), muchas de las cuales tienen un costo económico excesivo o demandan grandes conocimientos técnicos por parte de quienes desean investigar en esta área. Actualmente se están desarrollando códigos en R; no obstante, algunos de ellos continúan en periodo de prueba con el fin de corroborar sus estimaciones.

Finalmente, se recomienda promover estudios como este, con el fin de que los métodos psicométrico-cognitivos sean la base para la interpretación del desempeño en una prueba de rendimiento académico, sobre todo en cuanto a la finalización de la educación preuniversitaria, pues con ello es posible desarrollar programas de diagnóstico al ingresar a la educación superior, en los que se establezcan las habilidades consideradas como necesarias para proseguir estudios universitarios en conexión con las estructuras de conocimiento que se requieran para cursos posteriores. Esto es una necesidad, dado que actualmente se tienen evidencias del examen realizado por la Escuela de Matemática de la Universidad de Costa Rica, el cual está basado fundamentalmente en especificaciones de contenido, no en atributos cognitivos que reflejen de una manera más detallada el desempeño del estudiantado y, consecuentemente, permitan elaborar acciones que tiendan al mejoramiento de las deficiencias encontradas y a mantener las fortalezas que se evidencian en los resultados.

Además, una de las formas de validación de los resultados propios del *rule space* que no se han trabajado suficientemente en la literatura técnica es la utilización de análisis de funcionamiento diferencial de los ítems. Esta idea, propuesta por Dogan (2009), plantea básicamente que para cada ítem se defina el grupo de referencia como aquellas personas que dominan todos los atributos involucrados en el ítem y que el grupo focal esté constituido por quienes no dominen uno o más de los atributos involucrados en la resolución del ítem.

Es relevante tomar en consideración que el sistema educativo costarricense está viviendo una coyuntura debido a una reforma a los programas de estudio en Matemática. Su diseño en unidades o áreas y el marcado énfasis en la resolución de problemas como enfoque metodológico predominante ofrecen posibilidades de elaborar relaciones entre atributos cognitivos y con ello desarrollar investigaciones evaluativas que permitan dar cuentas tanto del logro de habilidades como de las trayectorias de aprendizaje que evidencia el estudiantado, incluso brindar información diferenciada según variables contextuales (tipo de institución, región educativa, condición socioeconómica, por ejemplo) que permitan comprender mejor el desempeño académico en pruebas de Matemática. La utilización progresiva de estos enfoques alternativos para la evaluación de los aprendizajes puede guiar un cambio en el norte de la evaluación en otras asignaturas del currículo escolar.

Finalmente, la utilización de modelos cognitivos para guiar el diseño y el desarrollo de pruebas de gran escala en el ámbito educativo es uno de los campos más promisorios. La elaboración de ítems a partir de estos modelos genera mediciones más apropiadas de los conocimientos y las habilidades del estudiantado. Por tanto, la información diagnóstica que se brinde será más precisa y permitirá guiar la elaboración de material didáctico, el planeamiento de estrategias de enseñanza acordes con el desarrollo de habilidades cognitivas en un área específica de conocimientos, así como el establecimiento de políticas educativas innovadoras que vinculen los resultados en estas pruebas con las habilidades que se hacen imprescindibles para desarrollarse en la sociedad, el ámbito académico y laboral.

Instituto de Investigación en Educación
Unidad de Información y Referencia
UNIE

Bibliografía

Artavia-Medrano, 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.

Birenbaum, M. & Tatsuoka, K. (1982). On the Dimensionality of Achievement Test Data. *Journal of Educational Measurement, 19*(4), 259-266.

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.

Birenbaum, M. & Tatsuoka, K. (1987). Open-Ended Versus Multiple Choice Response Formats-It Does Make a Difference for Diagnostic Purposes. *Applied Psychological Measurement, 11*(4), 385-395.

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.

Birenbaum, M., Kelly, A. & Tatsuoka, K. (1993). *Toward a Stable Diagnostic Representation of Students' Error in Algebra.* (Technical Report RR-92-58-ONR). Nueva Jersey: Educational Testing Service.

Birenbaum, M., Tatsuoka, C. & Yamada, T. (2004). Diagnostic Assessment in TIMSS-R: Between-Countries and Within-Countries Comparisons of Eighth Graders' Mathematics Performance. *Studies in Educational Evaluation, 30,* 151-173.

Birenbaum, M., Tatsuoka, K. & Gutvirtz, Y. (1992). Effects of Response Format on Diagnostic Assessment of Scholastic Achievement. *Applied Psychological Measurement, 16*(4), 353-363.

Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Borsboom, D. & Mellenbergh, G. (2007). Test Validity in Cognitive Assessment. En J. P. Leighton & M. J. Gierl (Eds.) *Cognitive Diagnostic Assessment: Theory and Applications* (pp. 85-115). Nueva York: Cambridge University Press.

Branch, J. (2000). *The trouble with think alouds: generating data using current verbal protocols.* Alberta, Canadá. Recuperado el 28 de abril del 2008 de http://www.slis.ualberta.ca/cais200/branch.htm

Brown, J. S. & Burton, R. R. (1977). *Diagnostic Models for Procedural Bugs in Basic Mathematics Skills*. (Reporte No. BBN-3669). San Diego, California. (Servicio de Reproducción de Documentos ERIC No. ED 159 036).

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 substraction: a cognitive perspective* (pp. 117-135). Hillsdale, Nueva Jersey: Erlbaum.

Bruning, R. H., Schraw, G. J. & Ronning, R. R. (2009). *Psicología cognitiva e instrucción*. Madrid: Alianza Editorial.

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*. (Technical Report RR-98-23). Princeton, NJ: Educational Testing Service.

Calfee, R. (1981). Cognitive Psychology and Educational Practice. *Review of Research in Education, 9*(3), 3-73.

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., MacDonald, G. & Leu, Y. (2011). Validating Cognitive Sources of Mathematics Item Difficulty. *The International Journal of Educational and Psychological Assessment, 7*(2), 74-93.

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*. San Francisco: Artículo presentado en Annual Conference of American Educational Research Association.

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 el 18 de marzo del 2012, de http://www.ierinstitute.org/IERI_Monograph_Volume_01.pdf

Chi, M. & Glaser, R. (1986). Capacidad de resolución de problemas. En R. Sternberg (Ed.) *Las capacidades humanas: un enfoque desde el procesamiento de la información* (pp. 293-324). Barcelona: Editorial Labor.

Cifarelli, V. (1993, abril). *Representation Processes in Mathematical Problem Solving*. (Reporte No. SE-053-900). California: Universidad de California en San Diego. Artículo presentado en Annual Meeting of the American Educational Research Association. (Servicio de Reproducción de Documentos ERIC No. ED 365 522).

Clements, D., Sarama, J., Tatsuoka, C. & Tatsuoka, K. (2013). *From Rasch Models to Rule Space and Poset-based Adaptive Testing: The Comprehensive Research-based Early Math Ability Test (CREMAT)*. Disponible en http://cadrek12.org/projects/using-rule-space-and-poset-based-adaptive-testing-methodologies-identify-ability-patterns-e

Cohen, L., Manion, L. & Morrison, K. (2008). *Research Methods in Education*. Nueva York: Routledge.

Coll, C. & Onrubia, J. (2001). Inteligencia, inteligencias y capacidad de aprendizaje. En C. Coll, J. Palacios & A. Marchesi (Comps.). *Desarrollo psicológico y educación: Psicología de la educación escolar* (pp. 189-210). Madrid: Alianza Editorial.

Conrad, F., Blair, J. & Tracy, E. (2002). *Verbal Reports are Data! A Theoretical Approach to Cognitive Interviews*. Proceedings of the Federal Committee on Statistical Methodology Research Conference. Washington: Office of Survey Methods Research. Recuperado el 4 de mayo del 2008 de http://www.bls.gov/ore/st/st990240.htm

Corter, J. & Tatsuoka, K. (2002). *Cognitive and Measurement Foundations of Diagnostic Assessments in Mathematics*. Technical Report of the College Board.

Crandall, B., Klein, G. & Hoffman, R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Cambridge, MA: The MIT Press.

De Brito, D. (2006). La verbalización como registro para análisis en la investigación sobre lectura. *Anales de Documentación, 9*, 43-51.

Dean, M. (2006). *Explaining performance in the Third International Math and Science Study (TIMSS) 1995 Advanced Mathematics Test*. Disertación doctoral no publicada, Columbia University, Nueva York, EE. UU.

Desimone, L. & Le Floch, K. (2004). Are We Asking the Right Questions? Using Cognitive Interviews to Improve Surveys in Education Research. *Educational Evaluation an Policy Analysis, 26*(1), 1-22.

DiBello, L. V., Stout, W. F. & Roussos, L. A. (1995). Unified Cognitive/Psychometric Diagnostic Assessment Likelihood-Based Classification Techniques. En P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.) *Cognitively Diagnostic Assessment*. Nueva Jersey: Lawrence Erlbaum Associates.

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, Nueva York, EE. UU.

Dogan, E. & Tatsuoka, K. (2008). An international comparison using a diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R. *Educ Stud Math, 68*, 263-272.

Embretson, S. & Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement, 38*(4), 343-368.

Ericcson, K. & Simon, H. (1993). *Protocol Analysis: Verbal Reports as Data.* Massachusetts: The MIT Press.

Eysenck, M. & Keane, M. (2005). *Cognitive Psychology: A Student's Handbook.* Nueva York: Psychology Press.

Ferrara, S. (2006). Toward a Psychology of Large-Scale Educational Achievement Testing: Some Features and Capabilities. *Educational Measurement: Issues and Practice, 25*(4), 2-5.

Ferrer, G. (2006). *Sistemas de evaluación de aprendizajes en América Latina: balance y desafíos.* Santiago: Alfabeta Artes Gráficas.

Fischer, G. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*, 359-374.

Fischer, G. (1997). Unidimensional Linear Logistic Rasch Models. En W. J. van der Linden & R. K. Hambleton (Eds.) *Handbook of Modern Item Response Theory* (pp. 225-243). Nueva York: Springer.

Frederiksen, N. (1990). Introduction. En N. Frederiksen, R. Glaser, A. Lesgold & M. Shafto (Eds.) *Diagnostic monitoring of skills and knowledge acquisition* (p. ix-xvii). Nueva Jersey: Erlbaum.

Galibert, M., Picón, J., Lozzia, G., Aguerri, M. & Attorresi, H. (2010). Componentes de dificultad de ítems para la evaluación de operaciones lógicas. Una aplicación del modelo LLTM. *SUMMA Psicológica UST, 7*(1), 3-14.

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. (1996). *An Investigation of the Cognitive Foundation Underlying the Rule-Space Model.* Disertación doctoral no publicada, University of Illinois at Urbana-Champaign, EE. UU.

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. & Zhou, J. (2008). Computer Adaptive-Attribute Testing. A New Approach to Cognitive Diagnostic Assessment. *Zeitschrift für Psychologie, 216*(1), 29-39.

Gierl, M. J., Leighton, J. P. & Hunka, S. M. (2007). Using the Attribute Hierarchy Method to Make Diagnostic Inferences About Examinees' Cognitive Skills. En J. P. Leighton &

Instituto de Investigación en Educación
Unidad de Información y Referencia

M. J. Gierl (Eds.) *Cognitive Diagnostic Assessment: Theory and Applications* (pp. 242-274). Nueva York: Cambridge University Press.

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., Leighton, J., Wang, C., Zhou, J., Gokiert, R. & Tan, A. (2009). *Validating Cognitive Models of Task Performance in Algebra on the SAT*. Nueva York: The College Board.

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.

Gierl, M., Zheng, Y. & Cui, Y. (2008). Using the attribute hierarchy method to identify and interpret cognitive skills that produce group differences. *Journal of Educational Measurement, 45*(1), 65-89.

Gitomer, D. H. & Yamamoto, K. (1991). Performance modeling that integrates latent trait and class theory. *Journal of Educational Measurement, 28*(2), 173-189.

Gorin, J. (2006). Test design with cognition in mind. *Educational Measurement: Issues and Practice, 25*(4), 21-35.

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, Nueva York, EE. UU.

Holyoak, K. J. (1995). Problem solving. En D. Osherson & E. E. Smith (Eds.), *Invitation to cognitive science*, Vol. 3: Thinking (2nd. ed) (pp. 267-296). Cambridge, MA: MIT Press.

Huff, K. & Goodman, D. (2007). The Demand for Cognitive Diagnostic Assessment. En J. P. Leighton & M. J. Gierl (Eds.) *Cognitive Diagnostic Assessment: Theory and Applications* (pp. 19-60). Nueva York: Cambridge University Press.

Hunka, S., Gierl, M. & Leighton, J. (2004). *Exploring the Logic of Tatsuoka's Attribute Rule-Space Model for Test Development and Analysis: A Tutorial.* Disponible en http://www.education.ualberta.ca/educ/psych/crame/files/Rulspc.nb

Im, S. (2007). *Statistical Consequences of Attribute Misspecification in the Rule Space Model.* Disertación doctoral no publicada, Columbia University, Nueva York, EE. UU.

Im, S. & Yin, Y. (2009). Diagnosing skills of statistical hypothesis testing using the Rule Space Method. *Studies in Educational Evaluation, 35*, 193-199.

Johnson, B. & Christensen, L. (2008). *Educational research: quantitative, qualitative, and mixed approaches*. Nueva York: Sage Publications.

Johnstone, C., Bottsford-Miller, N. & Thompson, S. (2006). *Using the think aloud method (cognitive labs) to evaluate test design for students with disabilities and English language learners* (Technical Report 44). Minneapolis: Universidad de Minnesota, National Center on Educational Outcomes.

Junker, B. & Sijtsma, K. (2001). Cognitive Assessment Models with Few Assumptions, and Connections with Nonparametric IRT. *Applied Psychological Measurement, 25*(3), 211-220.

Kasai, M. (1997). *Application of the rule space model to the reading comprehension section of the test of English as a foreign language (TOEFL).* Disertación doctoral no publicada, University of Illinois at Urbana-Champaign, 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). Nueva Jersey: Educational Testing Service.

Kerlinger, F. & Lee, H. (2002). *Investigación del comportamiento.* México: McGraw-Hill Interamericana.

Klahr, D. (1989). Information-Processing Approaches. *Annals of Child Development, 6*, 133-185.

Klahr, D. (2000). *Exploring Science: The Cognition and Development of Discovery Processes.* Cambridge: MIT Press.

Klahr, D. & Simon, H. A. (1999). Studies of Scientific Discovery: Complementary Approaches and Convergent Findings. *Psychological Bulletin, 125*(5), 524-543.

Lane, S. (1989). Implications of Cognitive Psychology for Measurement and Testing. *Educational Measurement: Issues and Practice, 8*(1), 28-30.

Leighton, J. (2004). Avoinding Misconception, Misuse, and Missed Opportunities: The Collection of Verbal Reports in Educational Achievement Testing. *Educational Measurement: Issues and Practice,* 23(4), 6-15.

Leighton, J. & Gierl, M. (2007a). Defining and Evaluating Models of Cognition Used in Educational Measurement to Make Inferences About Examineees' Thinking Processes. *Educational Measurement: Issues and Practice, 26*(2), 3-16.

Leighton, J. & Gierl, M. (2007b). Why Cognitive Diagnostic Assessment? En J. P. Leighton & M. J. Gierl (Eds.) *Cognitive Diagnostic Assessment: Theory and Applications* (pp. 3-18). Nueva York: Cambridge University Press.

Leighton, J. & Gierl, M. (2011). *The Learning Sciences in Educational Assessments: The Role of Cognitive Models.* Nueva York: Cambridge University Press.

Leighton, J., Cui, Y. & Cor, M. (2008, abril). *Testing Expert-Based and Student-Based Cognitive Models: An Application of the Attribute Hierarchy Method and Hierarchy Consistency*

Index. Artículo presentado en National Council on Measurement in Education, Nueva York, NY.

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). Recuperado el 9 de setiembre del 2007, de http://www.education.ualberta.ca/educ/psych/crame

Leighton, J., Gierl, M. & Hunka, S. (2004). The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement, 41*(3), 205-236.

Levine, M. & Rubin, D. (1979). Measuring the Appropriateness of Multiple-Choice Test Scores. *Journal of Educational Statistics, 4*(4), 269-290.

Li, J. & Klahr, D. (2006). The Psychology of Scientific Thinking: Implications for Science Teaching and Learning. En J. Rhoton & P. Shane (Eds.). *Teaching Science in the 21st Century*. National Science Teachers Association and National Science Education Leadership Association: NSTA Press.

Linacre, M. (2012b). WINSTEPS (Versión 3.74.0) [Software de cómputo]. Chicago, IL, EE.UU.: Winsteps Rasch Measurement.

Lohman, D. (2000). Complex information processing and intelligence. En R. J. Sternberg (Ed.) *Handbook of Intelligence*. (pp. 285-340). Nueva York: Cambridge University Press.

Mayer, R. (1986). Capacidad matemática. En R. Sternberg (Ed.) *Las capacidades humanas: un enfoque desde el procesamiento de la información* (p. 165-194). Barcelona: Editorial Labor.

McGlohen, K. (2004). *The Application of a Cognitive Diagnosis Model via an Analysis of a Large-Scale Assessment and a Computerized Adaptive Testing Administration*. Disertación doctoral no publicada, University of Texas at Austin, Texas, EE. UU.

Meijer, R. & Sijtmsa, K. (2001). Methodology Review: Evaluating Person Fit. *Applied Psychological Measurement, 25*(2), 107-135.

Messick, S. (1984). The Psychology of Educational Measurement. *Journal of Educational Measurement, 21*(3), 215-237.

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research, 45*, 35-44.

Milewski, G. & Baron, P. (2002, abril). *Extending DIF Methods To Inform Aggregate Reports on Cognitive Skills*. Nueva Orleans: Artículo presentado en Annual Meeting of the National Council on Measurement in Education. (Servicio de Reproducción de Documentos ERIC No. ED 466 712).

Mislevy, R. (1994). Evidence and inference in educational assessment. *Psychometrika, 59*, 439-483.

Mislevy, R., Steinberg, L. & Almond, R. (2003). On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 3-62.

Montague, M. & Applegate, B. (1993). Middle School Students' Mathematical Problem Solving: An Analysis of Think-Aloud Protocols. *Learning Disability Quarterly, 16*(1), 19-32.

Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Ediciones Pirámide.

Muñiz, J. (2002). *Teoría Clásica de los Tests*. Madrid: Ediciones Pirámide.

Muñiz, J., Fidalgo, A., García-Cueto, E., Martínez, R. & Moreno, R. (2005). *Análisis de los ítems*. Madrid: Editorial La Muralla.

National Research Council, NRC. (2001). *Knowing what students know: the science and design of educational assessment*. Washington: National Academy Press.

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.

Nokes, T. J., Schunn, C. D. & Chi, M. T. H. (2010). Problem Solving and Human Expertise. En P. Peterson, E. Baker & B. McGaw (Eds.). *International Encyclopedia of Education*. Vol. 5, pp. 265-272. Oxford: Elsevier.

Norris, S., Leighton, J. & Phillips, L. (2004). What is at stake in knowing the content and capabilities of children's minds?: A case for basing high stakes tests on cognitive models. *Theory and Research in Education, 2*(3), 283-308.

Norris, S., Macnab, J. & Phillips, L. (2007). Cognitive Modeling of Performance on Diagnostic Achievement Tests. En J. P. Leighton & M. J. Gierl (Eds.) *Cognitive Diagnostic Assessment: Theory and Applications* (pp. 61-84). Nueva York: Cambridge University Press.

Novick, L. & Bassok, M. (2005). Problem Solving. En K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 321-349). Nueva York: Cambridge University Press.

Onrubia, J., Rochera, M. J. & Barberà, E. (2001). La enseñanza y el aprendizaje de las matemáticas: una perspectiva psicológica. En C. Coll, J. Palacios & A. Marchesi (Comps.). *Desarrollo psicológico y educación: Psicología de la educación escolar* (pp. 487-508). Madrid: Alianza Editorial.

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.

Resnick, L. & Ford, W. (1990). *La enseñanza de las matemáticas y sus fundamentos psicológicos.* Barcelona: Ediciones Paidós Ibérica.

Rico, L., Marín, A. & Romero, I. (2001). Fines de la educación matemática y proyectos curriculares. En L. Rico (Ed.) *Bases teóricas del currículo de Matemáticas en Educación Secundaria* (pp. 319-375). Madrid: Editorial Síntesis.

Roberts, M., Brito, C., Gotzmann, A., & Gierl, M. (2009, abril). *Development of Cognitive Models in Mathematics to Promote Diagnostic Inferences about Student Performance.* Artículo presentado en Annual meeting of the American Educational Research Association, San Diego, CA.

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

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.

Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A. & Templin, J. H. (2007). The Fusion Model Skills Diagnosis System. En J. P. Leighton & M. J. Gierl (Eds.) *Cognitive Diagnostic Assessment: Theory and Applications* (pp. 275-318). Nueva York: Cambridge University Press.

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.

Rupp, A. & Mislevy, R. (2007). Cognitive Foundations of Structured Item Response Models. En 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.

Saito, H. (1998). *Cognitive diagnostic perspectives of a second language reading test.* Disertación doctoral no publicada, University of Illinois at Urbana-Champaign.

Schoenfeld, A. (2007). What is Mathematical Proficiency and How Can It Be Assessed? En A. H. Schoenfeld (Ed.) *Assessing Mathematical Proficiency.* (pp. 59-74). Nueva York: Cambridge University Press.

Scott, H. (1998). *Cognitive Diagnostic Perspectives of a Second Language Reading Test.* Disertación doctoral no publicada, University of Illinois at Urbana-Champaign, EE. UU.

Sheehan, K. (1997). A tree-based approach to proficiency scaling and diagnostic assessment. *Journal of Educational Measurement, 34*(4), 333-352.

Sheehan, K., Tatsuoka, K. & Lewis, C. (1993). *A Diagnostic Classification Model for Document Processing Skills.* (Technical Report RR-93-39-ONR). Nueva Jersey: Educational Testing Service.

Shaw, D. S. (1986). *Effects of adaptive diagnostic testing on two types of computerized remediation.* Disertación doctoral no publicada, University of Illinois at Urbana-Champaign.

Simon, H. A. (1978). Information-processing theory of human problem solving. En W. K. Estes (Ed.) *Handbook of learning and cognitive processes,* Vol. V (pp. 271-295). Hillsdale, NJ: Erlbaum.

Smith, E. & Kosslyn, S. (2008). *Procesos cognitivos: modelos y bases neurales.* Madrid: Pearson Prentice Hall.

SPSS, Inc. (2007). SPSS 16.0 [Software de cómputo].

Sternberg, R. (1986). *Las capacidades humanas: un enfoque desde el procesamiento de la información.* Barcelona: Editorial Labor.

Sung-Hoon, K. (1990). *Statistical classification of item-response patterns into misconception groups in rule space.* Disertación doctoral no publicada, University of Illinois at Urbana-Champaign, EE. UU.

Svetina, D., Gorin, J., Tatsuoka, K. (2011). Defining and Comparing the Reading Comprehension Construct: A Cognitive-Psychometric Modeling Approach. *International Journal of Testing, 11*(1), 1-23.

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. (1984). Caution Indices Based On Item Response Theory. Psychometrika, *49*(1), 95-110.

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 sensivity 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. (1991). *Boolean algebra applied to determination of the universal set of misconception states* (ONR-Technical Report No. RR-91-44). Princeton, NJ: Educational Testing Services.

Tatsuoka, K. (1995). Architecture of Knowledge Structures and Cognitive Diagnosis: A Statistical Pattern Recognition and Classification Approach. En P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.) *Cognitively Diagnostic Assessment* (pp. 327-359). Hillsdale, NJ: Erlbaum.

Tatsuoka, K. (1996). Use of Generalized Person-Fit Indexes, Zetas for Statistical Pattern Classification. *Applied Measurement in Education, 9*(1), 65-75.

Tatsuoka, K. (2009). *Cognitive Assessment: An Introduction to the Rule Space Method.* Nueva York: Routledge Taylor & Francis Group.

Tatsuoka, K. & Boodoo, G. (2000). Subgroup differences on the GRE Quantitative Test based on the underlying cognitive processes and knowledge. En A. E. Kelly & R. A. Lesh (Eds.), *Handbook of Research Design in Mathematics and Science Education* (pp. 821-858). Mahwah, NJ: Erlbaum.

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. (1982). Detection of Aberrant Response Patterns and Their Effect on Dimensionality. *Journal of Educational Statistics, 7*(3), 215-231.

Tatsuoka, K. & Tatsuoka, M. (1983). Spotting Erroneous Rules of Operation by the Individual Consistency Index. *Journal of Educational Measurement, 20*(3), 221-230.

Tatsuoka, K. & Tatsuoka, M. (1985). *Bug Distribution and Pattern Classification.* (Informe No. CERL-RR-85-3-ONR). Urbana, IL, EE. UU.: Illinois University, Urbana. Computer-Based Education Research Laboratory. (No. de servicio de reproducción de documentos ERIC ED 263 147).

Tatsuoka, K., Birenbaum, M. & Arnold, J. (1989). On the Stability of Students' Rules of Operation for Solving Arithmetic Problems. *Journal of Educational Measurement, 26*(4), 351-361.

Tatsuoka, K., Varadi, F. & Tatsuoka, C. (2004). PMAIN-RULE SPACE. [Software de cómputo no publicado]. Trenton, NJ, EE.UU.: Tanar.

Tatsuoka, K., Birenbaum, M., Lewis, C. & Sheehan, K. (1993). *Proficiency scaling based on conditional probability functions for attributes.* (Technical Report RR-93-50-ONR). Nueva Jersey: Educational Testing Service.

Tatsuoka, K., Corter, J. & Tatsuoka, C. (2004). Patterns of Diagnosed Mathematical Content and Process Skills in TIMSS-R Across a Sample of 20 Countries. *American Educational Research Journal, 41*(4), 901-926.

Tatsuoka, K., Guerrero, A., Corter, J., Tatsuoka, C., Yamada, T., Xin, T., et al. (2006). International comparisons of Mathematical Thinking Skills in the TIMSS-R. *Japanese Journal of Research on Testing*, 2(1), 3-40.

The R Foundation for Statistical Computing. (2013). R 3.0.1 [Software de cómputo].

van Someren, M., Barnard, Y. & Sandberg, J. (1994). *The think aloud method: a practical guide to modelling cognitive processes*. Amsterdam: Academic Press. Recuperado el 6 de abril del 2008, de http://staff.science.uva.nl/~maarten/Think-aloud-method.pdf

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. En M. I. Posner (Ed.), *Foundations of Cognitive Science* (pp. 526-579). Cambridge, MA: M.I.T. Press.

Whitely, S. E. (1981). Measuring Aptitude Processes with Multicomponent Latent Trait Models. *Journal of Educational Measurement, 18*(2), 67-84.

Wolfe, R. (2001). El dilema de la "granularidad" en el diseño del sistema de evaluación: cobertura curricular vs. cobertura poblacional. En P. Ravela (Ed.) *Los próximos pasos: ¿cómo avanzar en la evaluación de aprendizajes en América Latina?* Santiago: PREAL.

Xin, T., Xu, Z. & Tatsuoka, K. (2004). Linkage Between Teacher Quality, Student Achievement, and Cognitive Skills: A Rule-Space Model. *Studies in Educational Evaluation, 30*, 205-223.

Yepes-Baraya, M., Tatsuoka, K., Allen, N., O'Sullivan, L. & Hui, X. (1998, abril). *Application of rule space methodology to the 1996 NAEP science assessment: Grade 4 preliminary results.* Artículo presentado en National Council on Measurement in Education, San Diego, CA.

Zhou, J. (2010). *Estimating Attribute-Based Reliability in Cognitive Diagnostic Assessment.* Disertación doctoral no publicada, University of Alberta, Alberta, Canadá.

Zimowski, M., Muraki, E., Mislevy, R. & Bock, D. (2003). BILOG-MG (Versión 3.0) [Software de cómputo]. Lincolnwood, IL, EE.UU.: Scientific Software International, Inc.

Instituto de Investigación en Educación
Unidad de Información y Referencia

Anexo

FÓRMULA DE CONSENTIMIENTO INFORMADO[1]
(para ser sujeto de investigación)

Nombre del participante:_____

A. **PROPÓSITO DEL PROYECTO**: A partir de la utilización de los resultados de una prueba de Matemática de gran escala (aplicación nacional), se pretende identificar y evaluar las habilidades cognitivas que el estudiantado utiliza en la resolución de problemas. Esto con el propósito de que tanto estudiantes, como docentes, directores y directoras, padres, madres y personas encargadas tengan información que les sea de utilidad en cuanto a las fortalezas y debilidades que tienen el estudiantado y que se evidencian en el desempeño en una prueba escrita. El proyecto constituye la tesis de estudios doctorales del Programa Latinoamericano en Educación de la Universidad de Costa Rica, el cual se encuentra en la fase de recolección de datos, según la metodología propuesta para el estudio. Específicamente se espera obtener la manera en que el estudiantado resuelve los ítems de una prueba, mediante las verbalizaciones que hagan en el momento de la resolución, pues de esta manera se tendrá acceso a los conocimientos y las habilidades involucrados en dicha resolución a partir de quienes están involucrados en el aprendizaje del dominio específico (Matemática en este caso).

B. **¿QUÉ SE HARÁ?**: Se requiere la participación de seis estudiantes que cumplan las siguientes condiciones:1) cursar actualmente undécimo año, y 2) haber cursado el Ciclo Diversificado en el sistema educativo costarricense. Se espera que la selección esté a cargo de la docente de Matemática, quien a su vez deberá considerar una participación equilibrada entre sexos, así como una valoración de su desempeño en la asignatura que le permita expresar oralmente la forma en la que resuelve ejercicios. Cada persona seleccionada resolverá 60 ítems de selección única que corresponden a una prueba de bachillerato en Matemática administrada hace algunos años a nivel nacional. Deberán dejar constancia escrita del procedimiento utilizado en la resolución e ir verbalizando los conocimientos, las habilidades y los procedimientos que utilizó en la resolución de cada ítem. El trabajo será grabado en audio y video. En este último caso, no se enfocará la cara de la persona, sino únicamente el área correspondiente a la hoja en que resuelva el ítem. Las grabaciones de audio serán transcritas por el investigador. En ambos casos, las grabaciones se podrán utilizar tanto para el presente estudio como para otros con propósitos de avances en el conocimiento de esta área de trabajo. Las sesiones tendrán una duración aproximada de 2 h y 30 min y pueden extenderse según lo que demore cada estudiante en la resolución de la prueba. Se llevarán a cabo en una sala de la institución educativa a la que pertenezcan quienes hayan sido seleccionados y muestren disposición para participar.

---

[1]Este formulario está disponible en http://www.vinv.ucr.ac.cr/docs/formularios/formula-de-consentimiento-informado-1.doc. Constituye parte de la normativa de la Universidad de Costa Rica para desarrollar investigaciones. Fue aprobado en Sesión del Comité Ético Científico el 4 de junio del 2008. Las adaptaciones hechas para el presente estudio son responsabilidad del investigador.

C. **RIESGOS:** El estudio no involucra riesgos por la participación ni física ni mentalmente. El haber sido seleccionado por la docente que les imparte Matemática es un primer criterio para minimizar la incomodidad o la ansiedad que pueda provocarle estar en las sesiones de trabajo.

D. **BENEFICIOS:** Como resultado de su participación en este estudio, no obtendrá ningún beneficio directo, sin embargo, es posible que el investigador aprenda más acerca de los conocimientos, las habilidades y los procedimientos que utilizan el estudiantado de undécimo año en la resolución de ítems en Matemática y este conocimiento beneficie a otras personas en el futuro.

E. Antes de dar su autorización para este estudio usted debe haber hablado con Álvaro Artavia Medrano, quien debe haber contestado satisfactoriamente todas sus preguntas. Si quisiera más información más adelante, puede obtenerla llamando a Álvaro Artavia Medrano al teléfono 8949-8104 o bien escribiéndole al correo alvaro.artavia@ucr.ac.cr. Además, puede consultar sobre los derechos de los Sujetos Participantes en Proyectos de Investigación a la Dirección de Regulación de Salud del Ministerio de Salud, al teléfono 2257-2090, de lunes a viernes de 8 a.m. a 4 p.m. Cualquier consulta adicional puede comunicarse a la Vicerrectoría de Investigación de la Universidad de Costa Rica *a los teléfonos 2511-4201 ó 2511-5839*, de lunes a viernes de 8 a.m. a 5 p.m.

F. En caso de solicitarlo, la dirección de la institución educativa recibirá una copia de esta fórmula firmada.

G. Su participación en este estudio es voluntaria. Tiene el derecho de negarse a participar o a discontinuar su participación en cualquier momento.

H. Su participación en este estudio es confidencial, los resultados podrían aparecer en una publicación científica o ser divulgados en una reunión científica pero de una manera anónima.

I. No perderá ningún derecho legal por firmar este documento.

Instituto de Investigación
en Educación
Unidad de Información
Tencia

CONSENTIMIENTO

He leído o se me ha leído, toda la información descrita en esta fórmula, antes de firmarla. Se me ha brindado la oportunidad de hacer preguntas y éstas han sido contestadas en forma adecuada. Por lo tanto, accedo a participar como sujeto de investigación en este estudio

Nombre, cédula y firma del sujeto (niños mayores de 12 años y adultos)	Fecha
Nombre, cédula y firma del testigo (director de la institución)	Fecha
Nombre, cédula y firma del investigador que solicita el consentimiento	Fecha
Nombre, cédula y firma del padre/madre/representante legal (menores de edad)	Fecha