

## LOGRO EDUCATIVO: PRUEBA ENLACE MÉXICO 2008

Esthela Salas Simental, Tatiana Nayeli Domínguez Mendoza, Rosa María Farfán Márquez  
 Centro de Investigación y Estudios Avanzados del IPN (México)  
 esalas@cinvestav.mx, tdominguez@cinvestav.mx, rfarfan@cinvestav.mx

**Resumen.** Diferentes corrientes teóricas han intentado conocer cuáles son los factores que inciden en los procesos educativos, cuál es su relevancia y cómo podrían ser modificadas con la finalidad de obtener una mejor calidad en la educación. Fortalecer esta corriente de investigación en México es imprescindible. El objetivo de este trabajo es, identificar factores individuales, familiares y escolares incidentes en el nivel de logro académico en matemáticas de estudiantes de tercer año de secundaria. Se analizarán los resultados del examen de matemáticas propuesto por la Evaluación Nacional de Logro Académico en Centros Escolares (ENLACE). También se estudiarán los resultados de los cuestionarios de contexto que se aplicaron a una muestra representativa de estudiantes a sus profesores y a sus padres. Para ello se hará uso de Minería de Datos con el objetivo de encontrar relaciones ocultas entre las variables, sacar conclusiones y generar conocimiento a partir de estas.

**Palabras clave:** evaluación, género, prueba enlace, minería de datos

**Abstract.** Different research streams have tried to find factors which affect the educational process, what their significances are and how they might be modified in order to improve the quality of education. To strengthen this line of research in Mexico is essential. The aim of this study is to identify individual, family and school agents influencing mathematical academic achievement in junior high school. The results of the mathematics test proposed by National Assessment of Academic Achievement in Schools (ENLACE) will be analyzed. The context tests that were applied to a representative sample of students and their teacher and parents will also be analyzed. To do this work Data mining will be used with the aim to find hidden relations between variables, make conclusions and generate knowledge from them.

**Key words:** assessment, gender, enlace test, data mining

### Introducción

Desde diferentes corrientes teóricas se ha intentado conocer cuáles son los factores que inciden en los procesos educativos, cuál es su relevancia y cómo podrían ser modificadas con la finalidad de obtener una mejor calidad en la educación. Fortalecer esta corriente de investigación en México es, por lo tanto, una tarea académica imprescindible si se pretende desarrollar políticas eficaces para la mejora de la calidad y la equidad educativa (Blanco, 2008). Por esta razón es que en la actualidad, la búsqueda de factores asociados al logro escolar ha sido intensa.

Sobre la base de datos obtenida de la aplicación de la prueba ENLACE 2008 (instrumento que la Secretaría de Educación Pública (SEP) pone a disposición de la sociedad mexicana, ante las exigencias actuales de rendición de cuentas, en donde se proporciona información a los estudiantes, a los padres de familia, a los docentes, a los directivos de las instituciones educativas y a la sociedad en general, respecto del logro académico de los alumnos del Sistema Educativo Nacional de México), realizada a estudiantes de educación básica a nivel nacional, se

analizaron los factores de logro académico de los alumnos de tercero de secundaria en el Distrito Federal utilizando técnicas de minería de datos y estadística.

### Antecedentes

A inicios de 2006 la Secretaría de Educación Pública, como respuesta a una demanda de construcción de diagnósticos acerca del logro educativo de los alumnos, consideró importante desarrollar una evaluación a los estudiantes de educación básica. Uno de los objetivos de éste instrumento es que permitiera identificar indicadores orientados a la intervención pedagógica que impactaran de manera positiva en el mejoramiento de la calidad educativa. Dentro de este marco, la Unidad de Planeación y Evaluación de Políticas Educativas (UPEPE) de la SEP, a través de la Dirección General de Evaluación de Políticas (DGEP), implementó la Evaluación Nacional de Logro Académico en Centros Escolares (ENLACE) y, año tras año, esta prueba -cuyas especificaciones siguen las competencias y contenidos establecidos en los planes y programas de estudio oficiales vigentes y se encuentra enmarcada en un enfoque constructivo- se ha aplicado con el propósito fundamental de medir el resultado del trabajo escolar en las asignaturas de Español, Matemáticas y, en el año 2008 Ciencias Naturales.

Los instrumentos de prueba se diseñaron como pruebas objetivos, es decir, se conformó con reactivos de opción múltiple. Éstos instrumentos únicamente permite la selección de una respuesta correcta, renunciando a la posibilidad de explorar algunos aspectos, por un lado, en Español, la producción oral y escrita, las prácticas de búsqueda de información en fuentes diversas y/o la construcción de opiniones y valoraciones; por el otro, en Matemática, la creación de unidades arbitrarias de medida, el uso de instrumentos de geometría y/o la creación y exploración de objetos tridimensionales. Asimismo, con el fin de establecer una mejor precisión, se añadieron cuestionarios de contexto dirigidos a una muestra específica de estudiantes, padres y directores con el objetivo de detectar factores externos al proceso de enseñanza y aprendizaje que influyan en el rendimiento académico de los escolares.

Como lo plantea en el Manual Técnico de ENLACE-2008, “ENLACE es un instrumento estandarizado, objetivo, de alcance nacional, diseñado para que los maestros, maestras, directivos, autoridades educativas, investigadores y escolares de todo el país, dispongan de una medida válida, objetiva y confiable, del estado actual del aprovechamiento académico de los estudiantes” (Manual Técnico, 2008, pp. 2); cuyo designio es la obtención de resultados que permitan la toma de decisiones asociadas con la mejora educativa, como ya se ha dicho, sobre la base de los cuales posteriormente se formulen e implementen políticas educativas.

## Marco teórico

Estudios como el de Blanco (2007) proponen un marco de referencia que permite establecer niveles concretos que se encuentran relacionados con el logro académico. Estos niveles abarcan desde características del estudiante, así como rasgos significativos del ambiente en donde se desenvuelve, hasta características de su núcleo familiar y escolar. Para entender aún mejor esta relación se presenta el siguiente esquema:

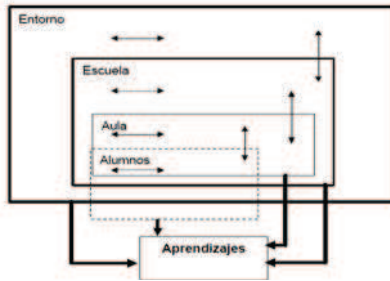


Figura 1. Niveles asociados a los resultados educativos (Blanco, 2007, pp.17).

Podemos inferir entonces, que la labor de identificar factores de logro académico dista de ser sencilla. Dada la complejidad de la búsqueda y teniendo en consideración los niveles anteriores, se tendrán como base para el análisis los siguientes estratos: en primer lugar, el nivel individual, comprendido por las características del alumno, su entorno y su familia; en segundo lugar, el nivel escolar, el cual considera las oportunidades de aprendizaje, clima en el aula, expectativas, política, participación, entre otras.

De lo anterior se desprende una amplia gama de variables por lo que es necesario utilizar métodos que nos brinden la mayor objetividad posible. Los métodos de estudio multinivel y jerárquicos (véase Oliver, Rosel y Jara, 2000; Blanco, 2007) parecen ser altamente apropiados para analizar este tipo de datos, ya que permiten determinar la influencia que ejercen, por separado, los distintos niveles involucrados en el fenómeno escolar. Sin embargo el auge de la tecnología y la computación así como su rápida implementación en actividades relacionadas con la educación han complementado los métodos estadísticos ya existentes, permitiendo una mayor eficacia en el procesamiento de datos. La existencia de algoritmos computacionales que pueden encontrar el significado de las palabras fue uno de los descubrimientos más excitantes en la ciencia cognitiva de las décadas pasadas. El KDD (Descubrimiento de Conocimiento en Bases de Datos) es un proceso no trivial de identificación válida, novedoso, potencialmente útil que encuentra patrones en datos (Coello, Dehuri y Ghosh, 2009). La Minería de Datos es una manera de desarrollar el KDD. Este método permite combinar métodos y herramientas de bases de datos, estadística e inteligencia artificial para tratar grandes cantidades de datos con el

objetivo de encontrar relaciones ocultas entre las variables, sacar conclusiones y generar conocimiento a partir de estas.

### **Datos**

(Las bases de datos trabajadas en este estudio fueron proporcionados por el Instituto Nacional de las Mujeres)

Población: 1 643 585 estudiantes de tercer año de secundaria en el año 2008 que contestaron la prueba ENLACE.

Dentro de las actividades en la aplicación de la prueba ENLACE, se hizo una selección de escuelas bajo un diseño muestral estratificado que estuvo a cargo de la Dirección General de Evaluación de Políticas (DGEP) de la Secretaría de Educación Pública (SEP), con el fin de aplicar cuestionarios de contexto tanto a estudiantes como a padres de familia y directivos. Cada estrato está descrito por la modalidad de la escuela: General, Particular, Técnica y Telesecundaria. Muestra: 173,691 estudiantes, los cuales según el género, el 52% fueron mujeres y el 48% hombres.

### **Instrumentos**

Para realizar el presente estudio se analizaron los resultados obtenidos de la asignatura de matemáticas, así como los datos de los cuestionarios de contexto que contestaron alumnos, padres y directivos. A continuación se describen brevemente las particularidades de los instrumentos de contexto.

El examen de matemáticas consta de 74 reactivos.

El cuestionario para alumnos consta de 142 preguntas, en el cual se abarcan los siguientes aspectos: Datos generales (sexo, edad, discapacidad, trabajo, tiempo libre), aspiraciones educativas, entorno familiar (características de sus padres, hermanos, maltrato), características de su vivienda, estudios de lengua extranjera y conocimientos de computación, hábitos de estudio, entorno y ambiente escolar, características de su escuela, características de sus profesores de español y matemáticas, relación con sus amigos.

En el cuestionario de padres se integraron 101 preguntas que abarcan los siguientes aspectos: características personales de la madre y del padre, situación laboral y económica, características de sus hijos, vivienda (Bienes y Servicios), apoyo a sus hijos (estudios), características de la escuela donde estudian sus hijos, características de los profesores de sus hijos.

El cuestionario aplicado a directivos contempla 134 preguntas que engloban los siguientes aspectos: datos personales (sexo, edad, escolaridad), experiencia laboral, infraestructura de la escuela (Bienes, servicios, entorno), características del personal docente y administrativo, gestión y organización escolar.

## Método

Como se mencionó anteriormente, en este trabajo se cuenta con una gran cantidad de variables categóricas, (142 para el caso estudiantes, 101 para el caso de padres y 134 para el caso de directivos), motivo por el cual se plantea el uso de diversos métodos para seleccionar y construir aglomeraciones de variables que facilite la interpretación de los resultados.

Inicialmente, se realizó un análisis multinivel preliminar; para después, aplicar el Algoritmo de Cluster (aglomeración) para estudiar las relaciones y comportamientos tanto de las variables como de los casos, entendiendo por casos, las respuestas individuales de cada uno de los alumnos encuestados en la muestra.

Se comenzó, con el análisis de las tablas de contingencia tomando como factor, la prueba Chi-cuadrada de Pearson, la cual parte de la hipótesis de que las variables son independientes. Para ser más precisos, se analizaron los datos con una confianza del 95% haciendo la asignación de  $\alpha = 0.05$  (lo cual corresponde al complemento porcentual de la confianza). El objetivo fue encontrar aquellas variables que son dependientes, para lograrlo se requirió refutar la hipótesis planteada. Por lo tanto, fue necesario encontrar aquellas variables cuyo valor de significación fuera menor a  $\alpha$ . Este primer análisis nos permitió estructurar grupos de variables con los cuales se trabajó posteriormente para hacer un análisis más preciso y certero.

## Resultados

La primera conclusión importante que se observa, luego del análisis de las frecuencias de las respuestas correctas en Matemáticas de cada uno de los ítems de la prueba ENLACE, es que en todos ellos los resultados de los hombres prevalecen sobre los de las mujeres, es decir, no sólo que la amplitud entre las frecuencias de los géneros es -la mayoría de las veces- muy amplia, sino que también en ningún momento la cantidad de respuestas acertadas por las mujeres es superior a la de los hombres. Se observó que aun siendo la cota superior del rango de calificaciones en las mujeres, mayor a la de los hombres, éstas poseen una concentración en las calificaciones centrales (entre 450 y 650), mientras que la distribución de los hombres es más homogénea.

### Distrito Federal

- Género: Femenino
- Población: 5540
- Rango de Calificaciones: 295-923
- Género: Masculino
- Población: 4928
- Rango de calificaciones: 288-910

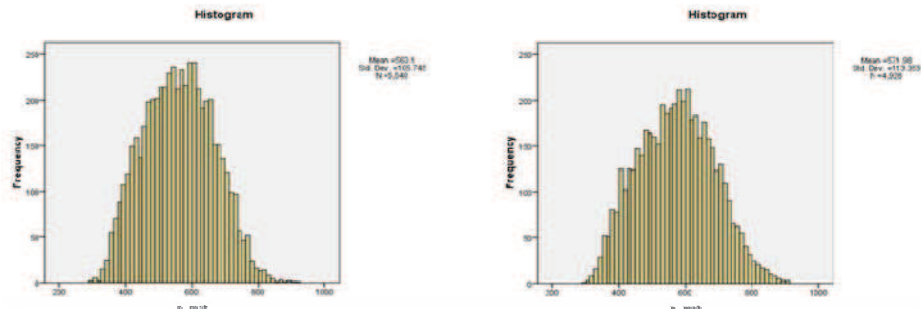


Figura 2. Histograma de puntaje de resultado de prueba de ENLACE en Matemática.

El primer experimento con el clúster jerárquico consistió en realizar dos corridas del algoritmo, una en la cual se incluyen únicamente los casos femeninos y otra donde se incluyen los casos Masculinos. En la siguiente tabla se muestran ocho categorías obtenidas después de la aplicación del algoritmo para los casos femeninos y masculinos. Se establece el rango de la puntuación para caracterizar cada una de las categorías.

Tabla I. Categorías por rango de calificaciones

Categoría	Mujeres		Hombres	
	Rango de Calificaciones	Número de Estudiantes	Rango de Calificaciones	Número de Estudiantes
1	295-389	245	288-375	152
2	388-451	633	376-463	760
3	449-541	1399	461-523	751
4	538-639	1671	521-586	960
5	637-745	1164	586-655	998
6	746-799	148	652-771	1040
7	802-852	38	772-839	146
8	863-923	8	840-910	41

En esta tabla se observa que los agrupamientos con mayor número de casos son los que se encuentran entre los rangos de puntuación matemática media y baja, siendo las mujeres los casos que mayor frecuencia tienen en las puntuaciones bajas. Después de analizar la tabla

anterior se realizó un análisis del comportamiento de cada una de las variables incluidas en el bloque de entorno familiar dentro de los agrupamientos antes definidos.

En la siguiente gráfica se observa el comportamiento de las variables cuya opción de respuesta fue siempre. Al parecer dentro de este bloque las variables “expresa afecto”, “platica de tus dudas” y “está al pendiente de tus calificaciones” están relacionadas con la puntuación en matemáticas. Sin embargo, las derivaciones de los gráficos indican que las diferencias en cuanto al apoyo de sus familiares a hombres y mujeres son relativamente pequeñas. Se observa que la diferencia de género más notoria es la relacionada con la variable “expresa de afecto” (Figura 3, colores rosas mujeres, colores oscuros hombres), ya que son las mujeres con mejor puntuación, las que expresan sentirse apreciadas por sus padres.

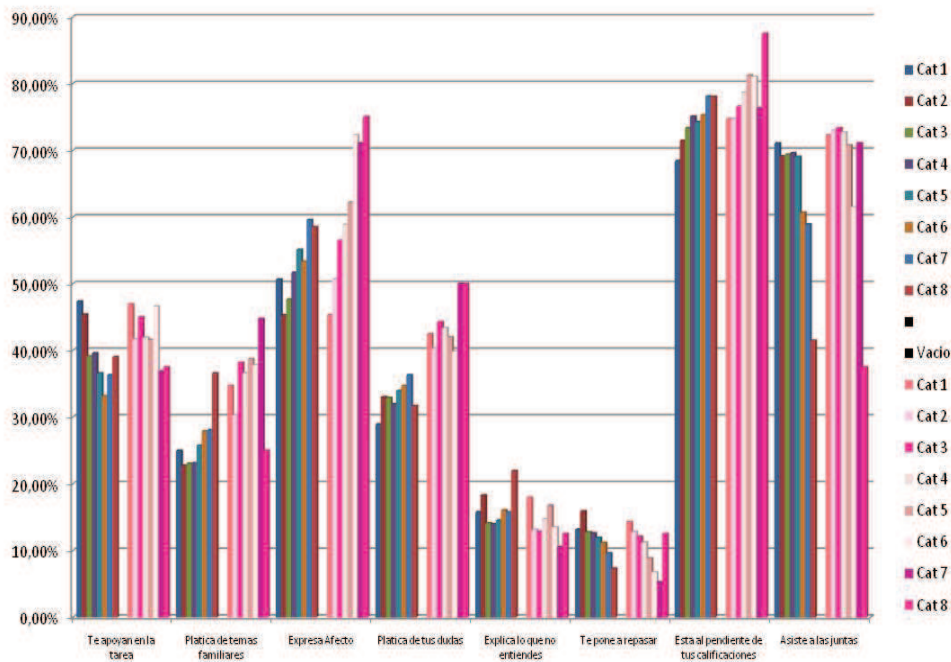


Figura 3. Análisis del clúster Entorno Familiar y Puntuación en Matemáticas.

El siguiente obstáculo planteado fue identificar qué conjunto de variables utilizar para realizar el agrupamiento de los casos, por lo cual, se empleó el mismo algoritmo de clúster jerárquico, para hacer agrupamiento de variables; las variables “El maestro de Matemáticas toma en cuenta las opiniones de los alumnos para mejorar la enseñanza” y “El maestro de Matemáticas me motiva para aprender más y seguir estudiando” son un ejemplo de estos agrupamientos que se relacionan con el desempeño de los estudiantes en matemáticas .

Otro resultado importante es que las variables “Pongo atención en las clases” y “¿Hasta qué nivel educativo te gustaría estudiar?” representa una influencia positiva (creciente) para el sistema de variables que se están analizando, donde la opción de respuesta siempre es la

categoría que más aporta a este comportamiento en la primer variable; mientras que para la segunda la opción que más aporta a este resultado es posgrado. Lo anterior da indicios de mejorar los modelos establecidos e indagar con mayor profundidad, la influencia de ciertas variables lo cual es una gran posibilidad de estudio.

### **Conclusiones**

Se observó que aun siendo la cota superior del rango de calificaciones en las mujeres, mayor a la de los hombres, éstas poseen una concentración en las calificaciones centrales (entre 450 y 650), mientras que la distribución de los hombres es más homogénea. Para enfatizar en el estudio de género, se requiere, por un lado, incluir en las pruebas de contexto ítems que nos permitan hacer inferencias más precisas y por el otro, incurrir en investigaciones sobre estudio de género que nos den indicios de una mejora en la elaboración de la prueba de conocimientos en matemáticas de la prueba ENLACE.

El entorno familiar influye en el buen desempeño del estudiante, a saber, en el caso de las mujeres se notó una diferencia respecto a los hombres cuando a éstas sus padres les demuestran afecto. Tanto en hombre como en mujeres es importante que los padres de familia estén al pendiente de las calificaciones de sus hijos.

Se observa que la mayoría de los hombres que tienen puntajes altos en la prueba de matemáticas expresan que su profesor de matemáticas les explica lo que no entienden. También se observó que es importante la organización que tienen las autoridades de la escuela.

### **Referencias bibliográficas**

- Blanco, E. (2007). Eficacia escolar en México. Factores escolares asociados a los aprendizajes en la educación primaria. Tesis de doctorado no publicada. Facultad Latinoamericana de Ciencias Sociales, FLACSO. México.
- Blanco, E. (2008). Factores escolares asociados a los aprendizajes en la educación primaria mexicana: Un análisis multinivel. REICE - Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 6(1), 58-84.
- Coello, C. A., Dehuri, S., Ghosh, S. (2009). Swarm Intelligence for Multi-Objective Problems in Data Mining. Berlin Heidelberg: Springer-Verlag.
- Manual Técnico ENLACE. (2008). Recuperado el 20 de noviembre de 2009 de <http://www.enlace.sep.gob.mx/ba/?p=manualtecnico>
- Oliver, Juan O., Rosel, J., Jara P. (2000). Modelos de regresión multinivel: aplicación en psicología escolar. Psicothema 12(3), 487-494.