

# Data Warehouse para analizar datos de pruebas educativas estandarizadas

Carolina Zambrano  
UDEC  
Chile  
carolinazambrano@gmail.com

Pedro Salcedo  
UDEC  
Chile  
paslagos@gmail.com

Dario Rojas  
UDEC  
Chile  
dfrojas@gmail.com

## ABSTRACT

In the educational sphere, much information comes from standardized tests. This information, when it is changing over time and large size, can be studied through multidimensional analysis provides by Data Warehouse system, making it possible to use this tool in the support decision making. This paper presents a data Warehouse Conceptual Schema that serves as a model to analyze and interpret the historical results of standardized tests, using as a case study the PISA test conducted by the OECD. Furthermore, the result of the analysis is related to public policies in the education area in the country Chile. From an educational standpoint the analyzes presented clearly show the benefit of multidimensional analysis, and they serve for historical tracking of the results obtained by PISA test of countries. All this is achieved by establishing a methodological framework to replicate the process carried out and can be applied in other different standardized educational tests.

## RESUMEN

En el ámbito educacional existe mucha información que proviene de fuentes como las pruebas estandarizadas. Esta información, al ser cambiante en el tiempo y de gran volumen, puede ser estudiada a través del análisis multidimensional que provee un Data Warehouse, haciendo posible utilizar dicha herramienta en el apoyo a la toma de decisiones. En este trabajo, se generó un Esquema Conceptual de Data Warehouse que sirve de modelo para analizar los resultados históricos de pruebas estandarizadas, utilizando como caso de estudio la prueba PISA realizada por la OECD. Además, el resultado del análisis es relacionado con las políticas públicas en el área educación en el país Chile. Desde un punto de vista educacional, los análisis presentados muestran claramente el beneficio del análisis multidimensional y sirven para realizar un seguimiento histórico de los resultados obtenidos por los países en la prueba PISA. Todo lo anterior se logra estableciendo un marco metodológico que permite replicar el proceso llevado a cabo con tal de ser aplicado en distintas pruebas educativas estandarizadas.

## Categories and Subject Descriptors

Education---Data management systems---Software Engineer

## General Terms

Education, Data Analysis

## Keywords

Multidimensional analysis, data warehouse, standardized tests.

## 1. INTRODUCCIÓN

En la actualidad, las organizaciones educativas cuentan con mucha información proveniente de procesos académicos, procesos de gestión curricular, entre otros. Estos datos suelen registrarse en sistemas de información. También existen fuentes de información educacional en los estudios que se realizan al interior de los establecimientos tal como; estudios sobre estilos de aprendizaje de los estudiantes, estudios sobre estrategias de aprendizaje tales como [9, 10, 12, 16, 24, 32]. Estos estudios deberían ser registrados de forma sistemática. Y de esta forma se podría utilizar esta información para realizar un seguimiento. Una forma de realizar este seguimiento es a través de un Data Warehouse que permite un estudio de tipo longitudinal.

En el ámbito internacional, la prueba estandarizada PISA (Programme for International Student Assessment) representa una fuente importante de información educacional. PISA es realizada por la OECD (Organisation for Economic Cooperation and Development). Otras pruebas estandarizadas internacionales son TIMSS (Trends in International Mathematics and Science Study) y PIRLS (Progress in International Reading Literacy Study). En este sentido, podemos considerar que realizar análisis sobre los datos de los resultados de pruebas estandarizadas es una oportunidad para encontrar respuestas desde los mismos datos y revisar cuál ha sido la tendencia histórica de los indicadores que se miden en el tiempo y que pueden ser explorados desde distintas dimensiones.

PISA está dirigida a estudiantes de 15 años [23] y se realiza cada tres años. En cada ciclo se enfatiza un dominio de evaluación. Los dominios son Lectura, Matemáticas y Ciencias. En el año 2000 el principal dominio fue Lectura, en el año 2003 el dominio era Matemáticas, en el año 2006 el dominio correspondió a Ciencias y así sucesivamente. En Chile se aplica desde el año 2000. El objetivo de PISA es comparar los resultados de aprendizaje de los estudiantes en los distintos sistemas educativos de los países miembros de la OECD.

En este contexto nos podemos preguntar, ¿es posible utilizar la información de PISA para revisar tendencias históricas de los estudiantes respecto de la variable socioeconómica y estudios de los padres usando un Data Warehouse? ¿Estos resultados se pueden relacionar con las políticas públicas en el área educación de cada país?

Desde el punto de vista de la tecnología cabe preguntarse, ¿qué ventajas nos ofrece el análisis multidimensional de los datos de PISA usando un Data Warehouse?, ¿es posible generar un esquema conceptual que sirva de modelo para interpretar los resultados históricos de pruebas estandarizadas usando los Data Warehouse?

Para dar respuesta a estas interrogantes se pueden utilizar herramientas de Minería de Datos Educativa tal como Data Warehouse [4,18] y/o minería de datos [2,15] aunque cabe precisar que en este estudio en curso sólo nos centraremos en los Data Warehouse como etapa inicial.

## 2. Data Warehouse

Según Chaudhuri [4] un Data Warehouse es una colección de datos orientada a un determinado ámbito, por ejemplo, al ámbito educativo. Además, es integrado y variable en el tiempo, ya que reúne información histórica (información cambiante en el tiempo). Y es utilizado para procesar información histórica a través del análisis multidimensional generalmente como apoyo a la toma de decisiones.

Los elementos principales de un Data Warehouse son dimensiones, hechos y medidas [4,17,18].

Una dimensión permite agrupar información respecto a un aspecto para el análisis [4,17,18]. Algunos ejemplos de dimensiones en el ámbito educativo son estudiantes, tipo de prueba que rinde un estudiante, situación socioeconómica a la que pertenece un estudiante, carrera que cursa, etc. Es importante señalar que la dimensión tiempo siempre está presente.

Los hechos representan un cruce de información entre dos o más dimensiones [4,17,18], por ejemplo, el cruce de información entre la dimensión estudiante y la dimensión tipo de prueba que rinde.

Las medidas son valores numéricos que resultan desde el cruce de las dimensiones y representan indicadores que dan utilidad a los hechos. Un ejemplo de medida es la “cantidad de estudiantes por cada tipo de prueba según su situación socioeconómica”, donde la medida es cantidad de estudiantes y las dimensiones cruzadas son tipo de prueba, estudiante y situación socioeconómica. En este sentido, un Data Warehouse es una herramienta que permite una rápida navegación y reportabilidad de las medidas y dimensiones de una forma más ágil y eficiente computacionalmente que realizar los mismos cálculos en una base de datos tradicional. Adicionalmente, se debe considerar que las medidas representan valores numéricos, así el tipo de análisis que se puede realizar con un Data Warehouse es de tipo cuantitativo.

En este sentido, el análisis de datos educativos mediante Data Warehouse y/o minería de datos, ha sido recalado por algunos autores que han indicado que las instituciones de educación superior pueden utilizar la información que se acumula de sus procesos curriculares, académicos, ambientes virtuales de aprendizaje, etc., con el fin de mejorar la gestión educativa [11,21]. De este modo, nace el concepto de Minería de Datos Educativa (MDE) [1,30] que otorga un sin número de ventajas y oportunidades, como lo señalan Calvet y Juan [3]. No obstante, lo anterior, deseamos destacar que ningún autor ha estudiado la utilidad de los Data Warehouse en datos educativos que provienen desde estudios como la prueba PISA o cualquier otro tipo de estudio educativo que provenga de pruebas estandarizadas.

En virtud de ello, podemos dar cuenta de investigaciones relacionadas con Data Warehouse en contextos educativos tales como los trabajos [19, 28, 36, 37, 38]. En estos trabajos, la fuente de información proviene de sistemas curriculares. Además, entre

las investigaciones que se han concentrado en aplicar técnicas de minería de datos a datos educativos sin utilizar Data Warehouse encontramos el trabajo [29] que aplica minería de datos sobre la información generada en el entorno virtual de aprendizaje Moodle y el trabajo de [26] que utiliza redes neuronales para predecir el éxito académico de estudiantes usando los datos de PISA.

En consecuencia, dadas las ventajas de aplicar el enfoque de MDE ya sea mediante Data Warehouse y/o minería de datos es que resulta interesante plantear este proyecto, de carácter preliminar, con el objetivo de realizar un análisis multidimensional de los datos educativos que provienen del estudio que se realiza a través de la prueba PISA. De este modo se podría usar el Data Warehouse como herramienta de apoyo al seguimiento de las políticas públicas, en el sentido que permite un estudio longitudinal.

La propuesta considera el diseño e implementación del Data Warehouse usando dos indicadores incluidos en PISA que son HISEI y HISCED. El primero corresponde al mayor índice socioeconómico de los padres (HISEI) y el segundo representa al mayor grado de educación alcanzado por los padres (HISCED). Estos índices fueron seleccionados debido a que están presentes en todas las versiones del estudio de PISA que se pueden descargar desde la página de la OECD.

Es importante indicar que el diseño e implementación del Data Warehouse usando datos de un estudio de gran escala como lo es PISA permitirá mostrar su utilidad en un caso que no es común, ya que la mayoría de las investigaciones han realizado análisis desde información proveniente de procesos académicos y no desde información proveniente de estudios, como PISA. Además, nuestro estudio considera la propuesta de un esquema conceptual que sirve como modelo para analizar los resultados históricos de pruebas estandarizadas usando como herramienta los Data Warehouse. Este modelo se esquematiza a continuación, en la metodología.

## 3. Metodología

La metodología se compone de 4 etapas. La primera etapa consiste en determinar dimensiones, medidas e indicadores desde datos de pruebas estandarizadas. Luego, en la segunda etapa se diseña el esquema conceptual del Data Warehouse para pruebas estandarizadas. La tercera etapa corresponde al análisis de información desde el Data Warehouse de pruebas estandarizadas. Y la última etapa es proyectar el análisis histórico obtenido desde el Data Warehouse como aporte a las políticas educativas del país en estudio.

La Figura 1 muestra un esquema que resume la metodología y sus etapas con el objetivo de establecer un esquema conceptual que sirva de soporte para analizar datos que provienen de pruebas estandarizadas como PISA usando como herramienta un Data Warehouse. Es importante resaltar que para validar la metodología se utiliza los datos de PISA.

Cabe mencionar que este estudio es de tipo longitudinal, y la muestra está dada por la cantidad de registros que se cargaran en la implementación del Data Warehouse.

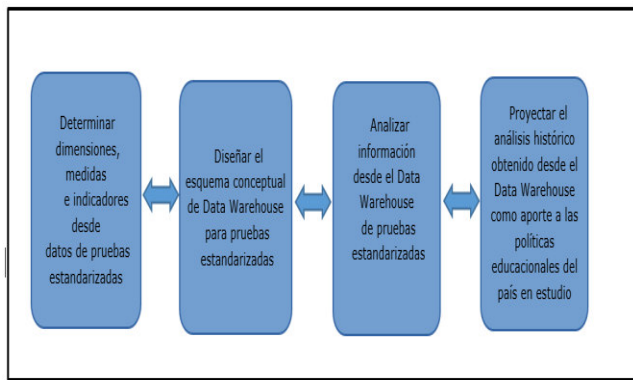


Figura 1: Esquema metodológico, con las etapas para analizar información desde pruebas estandarizadas usando un Data Warehouse.

### 3.1 Etapa 1: Determinar dimensiones, medidas e indicadores desde datos.

Esta es la primera etapa del marco metodológico presentado en la Figura 1 y consiste en determinar dimensiones, medidas e indicadores desde los datos de pruebas estandarizadas. En este estudio se utilizan los datos de PISA que son de libre disposición y se pueden descargar desde la página Web de la OECD bajo el nombre de Manual for the PISA database.

Tabla 1: Dimensiones del Data Warehouse PISA.

Dimensiones	Descripción de las Dimensiones
Alumno	Describe las características de alumno.
País	Representa las características de país como el nombre del país.
Tipo de Prueba	Indica el tipo de prueba Ciencia, Lectura, Matemática.
Socioeconómica	Indica el nivel socio económico
Tiempo	Indica los años de aplicación de PISA
Padres	Indica el nivel de escolaridad de los padres

La Tabla 1 y 2 muestran las dimensiones y medidas que se determinaron producto del análisis riguroso de las fuentes de PISA.

Tabla 2: Medidas del Data Warehouse PISA

Medidas e Indicadores	Descripción de las Medidas e Indicadores
Cantidad de estudiantes	Medida que sirve para mostrar la cantidad de estudiantes.
Promedio	Medida que sirve para mostrar el promedio de estudiantes.
HISEI	Indicador sobre el mayor índice socio económico de los padres.
HISCED	Indicador sobre el mayor grado de educación alcanzado por los padres del estudiante.

Es importante indicar que el tamaño de la muestra en este caso está representado por la cantidad de registros que se cargan en el Data Warehouse. La herramienta que se utiliza para implementar el Data Warehouse es SQL Server [14].

### 3.2 Etapa 2: Diseñar el esquema conceptual del Data Warehouse.

En esta segunda etapa del marco metodológico, presentado en la Figura 1, se diseñó el esquema conceptual del Data Warehouse de pruebas estandarizadas. Para realizar el diseño multidimensional de un Data Warehouse existen enfoques impulsados por los datos, enfoques impulsados por la demanda y enfoques híbridos [5]. En el caso de PISA se llevó a cabo un enfoque orientado a los datos, lo que consistió en analizar detalladamente las fuentes de datos de la prueba estandarizada PISA para determinar los elementos del Data Warehouse tales como medidas y dimensiones que eran posibles de considerar con los datos disponibles.

Es importante indicar que en el diseño de un Data Warehouse se distinguen tres etapas que son diseño conceptual, diseño lógico y diseño físico. Este último sirve para la implementación final antes de realizar el análisis multidimensional [22,25,31].

En esta investigación, el diseño conceptual del Data Warehouse representa el modelo de Data Warehouse para pruebas estandarizadas como PISA y es independiente de la implementación. Esta independencia ha sido estudiada y explicada en diversos trabajos [4,22,25]. La notación que se utiliza para representar el esquema conceptual del Data Warehouse para la prueba estandarizada PISA es la definida por Malinowski y Zimányi [22] en el modelo MultiDim.

La Figura 2 muestra el esquema conceptual del Data Warehouse para pruebas estandarizadas como PISA.

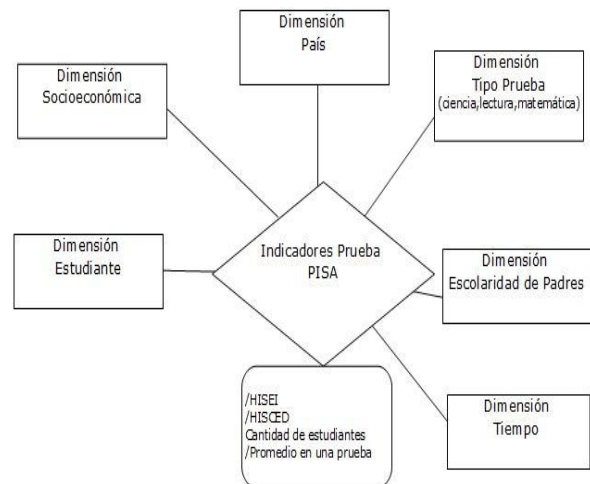


Figura 2: Esquema Conceptual del Data Warehouse para prueba estandarizada, usando notación de Malinowski y Zimányi [22].

El esquema conceptual de la Figura 2 tiene seis dimensiones, que representan las diferentes perspectivas desde las cuales se puede analizar la información de PISA.

### 3.3 Etapa 3 y 4: Analizar información desde Data Warehouse.

Posterior al diseño, se implementa el Data Warehouse para analizar datos de pruebas estandarizadas. Esto implica generar el esquema mostrado en la Figura 2 a nivel lógico y físico en la herramienta Sql Server. En ésta se cargan los registros históricos de la prueba estandarizada PISA para los años 2000, 2003, 2006 y 2009.

Una vez implementado el Data Warehouse, se utiliza el análisis multidimensional que consiste en obtener información resumida de tipo cuantitativa representada por tendencias históricas mediante el cruce de los datos a través de las dimensiones generadas en el Data Warehouse. Es importante indicar que la dimensión tiempo siempre estará presente en un análisis multidimensional usando Data Warehouse y permite representar las tendencias en el tiempo.

La proyección del análisis multidimensional en las políticas públicas se realiza revisando estudios de carácter cualitativo. De esta forma se triangula la información cuantitativa que otorga el Data Warehouse a través de las tendencias históricas con lo que se ha estudiado respecto a los modelos educativos de los distintos países.

En los resultados que se presentan en el siguiente apartado se utiliza el caso del país Chile para proyectar el análisis en las políticas públicas de dicho país.

## 4. Resultados

En este apartado se exponen una serie de reportes generados a través de la implementación del Data Warehouse para la prueba PISA.

La Figura 3 muestra la evolución de los puntajes promedio de los países participantes de Latinoamérica para los años 2000, 2003, 2006 y 2009. La tendencia muestra un crecimiento sostenido en Chile y Brasil logrando una mejora por sobre los 30 puntos cada uno.

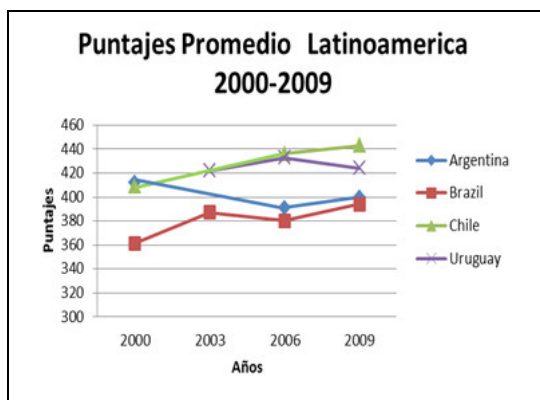


Figura 3 - Puntajes Promedio Latinoamérica 2000, 2003, 2006 y 2009.

La Figura 4 muestra los puntajes promedio agrupados por nivel socioeconómico en Chile para los años 2000, 2003, 2006 y 2009. Existe una diferencia de más de 100 puntos entre los niveles socioeconómicos más bajos y los más altos. Esto nos indica que el

nivel socioeconómico es factor que incide en los resultados como ya se ha demostrado en otros estudios [26,34].

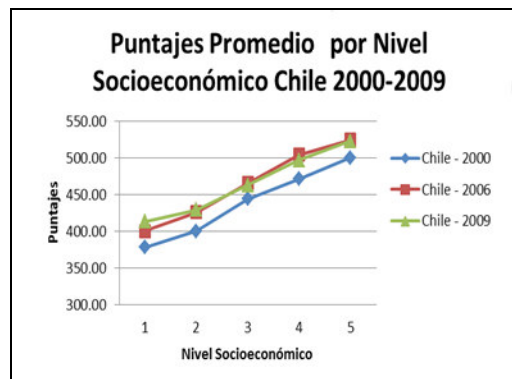


Figura 4 - Puntajes Promedio por Nivel Socioeconómico

La Figura 5 muestra los puntajes promedio agrupados por el nivel de escolaridad alcanzado por los padres de los estudiantes. Los niveles están representados desde 0 a 6 como se indica a continuación:

- Nivel 0: Sin Educación.
- Nivel 1: Primaria.
- Nivel 2: Primer ciclo de Secundaria.
- Nivel 3: Secundaria Humanista.
- Nivel 4: Secundaria Nivel Técnico.
- Nivel 5: Pregrado.
- Nivel 6: Postgrado.

Desde la información presentada en la Figura 5, se observa que en Chile existe relación entre el puntaje obtenido por el estudiante y el nivel de escolaridad de los padres. Si se relaciona el resultado mostrado en la Figura 5 con los resultados mostrados en la Figura 4, se aprecia el mismo patrón en la tendencia histórica. Este resultado refleja la agrupación que la sociedad chilena ha presentado en su sistema Educativo, donde, a mayores recursos económicos y mayor nivel de escolaridad de los padres, se evidencian mejores resultados de sus hijos en los puntajes obtenidos en pruebas como PISA. Esto también se refleja en estudios como el de Salazar y Leihy [33].

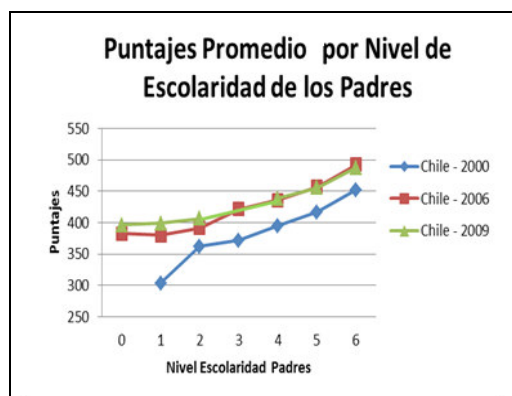


Figura 5 - Puntajes Promedios por Nivel de Escolaridad de los Padres

La Figura 6 muestra la evolución de los puntajes promedio de las distintas pruebas entre 2000, 2003, 2006 y 2009. Como se observa, la prueba de Lenguaje presenta el mayor progreso con una diferencia de 60 puntos entre el 2000 y el 2009, seguida de Ciencias con una mejora de 30 puntos. Sin embargo, en 2012 el puntaje promedio en Lenguaje fue de 441 lo que indica un estancamiento. Respuestas a esta situación se encuentran en el modelo educacional chileno y en las políticas públicas en educación como lo señalan también los siguientes estudios [7,8,34].

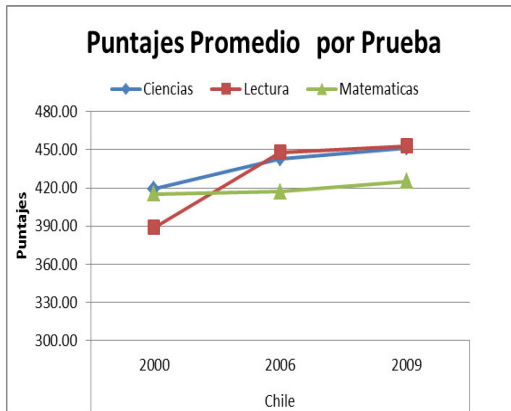


Figura 6 - Puntajes Promedio por Prueba para Chile

La Figura 7 muestra la evolución de los puntajes promedio para la prueba de Lenguaje por género entre 2000, 2003, 2006 y 2009 en Chile. Las mujeres mostraron una clara tendencia en alza en esta prueba, mientras que los hombres mantuvieron sus resultados entre el 2006 y el 2009.



Figura 7 - Puntajes Promedios de Lenguaje por Género

La Figura 8 muestra la tendencia de los puntajes promedio para la prueba de ciencias por género entre los años 2000, 2003, 2006 y 2009 en Chile. Tanto mujeres como hombres mostraron una tendencia positiva. Sin embargo, los hombres en promedio obtuvieron mejores resultados en esta prueba. Esta misma tendencia se observa en la Figura 9.



Figura 8 - Puntajes Promedio en Ciencias por Género

Se observa en la Figura 9 la evolución de los puntajes promedio para la prueba de matemáticas por género entre los años 2000, 2003, 2006 y 2009 en Chile. En esta prueba el puntaje de los hombres presentó una clara tendencia positiva, mientras que las mujeres no presentaron una tendencia positiva. La tendencia de los hombres con mejor puntaje en ciencias es un claro reflejo de la cantidad de mujeres que estudian carreras del área ciencias y matemática. Estudios relacionados con la presencia de mujeres en educación superior coinciden con esta tendencia de baja relación entre ciencia, matemática y mujeres [6, 27]. Por lo cual una tarea importante y un trabajo aún pendiente es fomentar la presencia de mujeres en ciencias, matemáticas e ingeniería [13,20].

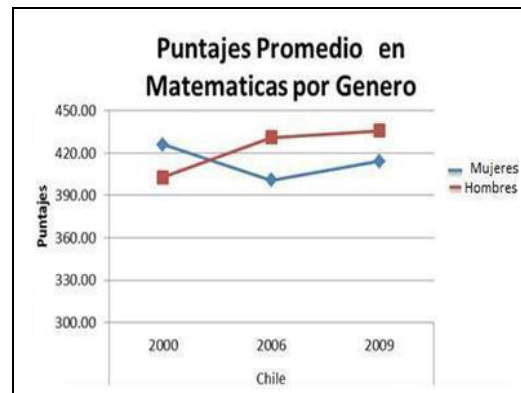


Figura 9 - Puntajes Promedios en Matemáticas por Género

La Figura 10 muestra los puntajes promedios de Chile y algunos países miembros de la OECD. Como se observa algunos países miembros de la OECD tuvieron una clara tendencia a la baja. Mientras que otros mantuvieron niveles de rendimiento. Es interesante observar que Chile se encontraba a 30 puntos de los países miembros de la OECD.

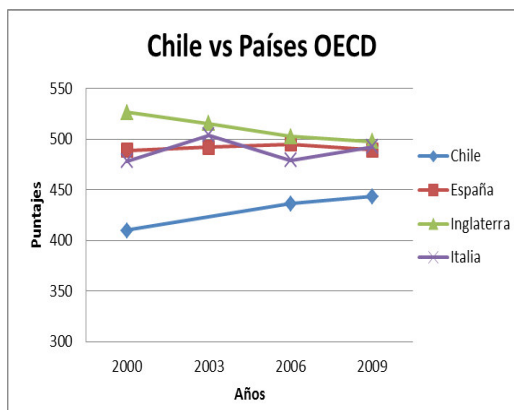


Figura 10 - Chile VS Países OECD

La Figura 11 muestra el nivel de escolaridad de los padres para los países de Latinoamérica para los años 2000, 2003, 2006 y 2009. Se observa que Chile junto con Argentina muestran los niveles de escolaridad más altos.

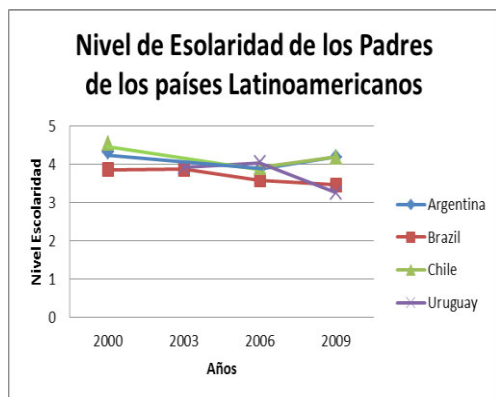


Figura 11 - Nivel de Escolaridad de los Padres en los países latinoamericanos

En síntesis, los resultados analizados anteriormente desde la información presentada en los gráficos, nos proporciona un primer acercamiento a la potencialidad del análisis multidimensional que nos otorga la tecnología de los Data Warehouse [4,17,18]. La potencialidad está en la rapidez que otorga la tecnología para procesar el volumen de datos que otorgan los estudios de gran escala con información de pruebas estandarizadas como PISA, TIMSS, PIRLS y SIMCE5. Es fundamental indicar que la tecnología que permite procesar la información del Data Warehouse implementado es OLAP (On-Line Analytical Processing) como lo señala Chaudhuri y Umeshwar [4]. Esta tecnología es la que permitirá potencialmente soportar la gran cantidad de datos históricos que se cargarán en el tiempo en el Data Warehouse diseñado. Es decir, cuando se agreguen los datos del estudio PISA 2012 y 2015 no será necesario redefinir el esquema conceptual pues éste ya ha sido concebido como un modelo para analizar los datos de PISA. Metodológicamente esta independencia entre el diseño conceptual y la implementación representan una gran ventaja.

Desde un punto de vista educacional los análisis presentados, a pesar de que son preliminares, muestran claramente el beneficio del análisis multidimensional. Y sirven para realizar un seguimiento histórico de los resultados obtenidos por los países en la prueba

estandarizada PISA. De este modo, los resultados se pueden relacionar con las políticas públicas en el área educación de los distintos países. Así las acciones que se llevan a cabo en dichas políticas públicas de cada país se verán reflejadas de forma positiva o no en los resultados de las mediciones estandarizadas como PISA.

También se aprecia que la información obtenida a través del análisis multidimensional se puede complementar y/o validar con lo que estudios de carácter cualitativo han indicado respecto a las políticas públicas en el ámbito educacional.

Finalmente, se puede señalar que los indicadores que se analizan multidimensionalmente a través de un seguimiento histórico, sirven como un suministro para realizar análisis en pro de tomar decisiones de forma científica. Es decir, si las tendencias históricas de los datos indican que existen problemas, entonces se debería actuar para buscar soluciones que permitan mejorar los resultados. Esto incidirá de forma positiva en la calidad de la educación.

A continuación, se presenta la discusión en relación a los resultados obtenidos, la funcionalidad y generalidad del esquema conceptual del Data Warehouse para pruebas estandarizadas y algunas relaciones de los resultados de Chile con las políticas públicas en educación y la proyección del trabajo realizado.

## 5. Discusión

De los resultados analizados con anterioridad es importante señalar que en el ámbito educacional es relevante estudiar el progreso histórico de los resultados obtenidos en estudios de carácter internacional como lo es PISA con el fin de evaluar si las acciones que cada gobierno impulsa a través de sus políticas públicas en esta área muestran resultados significativos en el tiempo.

En este sentido, los resultados del análisis multidimensional usando un Data Warehouse mostraron que Chile fue el país con los mejores puntajes promedio de Latinoamérica y que junto con Brasil, lograron mejoras más significativas para los años 2000, 2003, 2006 y 2009. También se mostró una relación directa entre el nivel socioeconómico de los estudiantes y su puntaje promedio alcanzado en las pruebas PISA. Lo que significa que a medida que aumentó el nivel socioeconómico del estudiante también lo hizo su puntaje promedio. Así coincidimos con [26] que en su trabajo utilizan redes neuronales para predecir el éxito académico de estudiantes usando los datos de PISA y logran determinar que la variable socioeconómica tiene un gran valor predictivo del rendimiento académico.

También se puede señalar que al analizar los resultados de Chile obtenidos a través del Data Warehouse de PISA existe una clara relación de ellos con las políticas públicas que dirigen el sentido de la gestión de la enseñanza obligatoria en Chile entre los periodos de 1990 y 2010. Así se observa que existe coincidencia con lo que señalan Donoso y Donoso [7, 8] respecto de que se ha logrado progresos importantes en materia de cobertura e incremento presupuestario, sin embargo, no han sido equivalentes sus avances en la calidad de la educación, pese a que en la comparación internacional PISA Chile mejora su posición.

En términos generales, el análisis multidimensional de los datos de PISA presentado también muestra que algunos países desarrollados han paulatinamente disminuido sus puntajes promedios en las

diferentes pruebas esto debería encontrar respuestas en las políticas públicas de cada país. Por lo cual el Data Warehouse de los datos de PISA es una herramienta útil para realizar un seguimiento histórico de las acciones o programas que se llevan a cabo en cada país.

Respecto de la utilidad de los Data Warehouse en el ámbito de la educación coincidimos con otros autores como [19, 28, 36, 37, 38] respecto a sus ventajas como herramienta para realizar un seguimiento en base a la información histórica que se almacena en el Data Warehouse. De este modo, el seguimiento debe apuntar a tomar acciones en pro de mejorar algún aspecto en el ámbito educacional.

Sobre el esquema conceptual que representa el modelo para el análisis de datos de la prueba estandarizada PISA es importante indicar que su sintaxis gráfica está representada por el modelo MultiDim de Malinowski y Zimányi [22]. Por esta razón, al ser un esquema conceptual, se hace independiente de su implementación, es decir, se podría utilizar cualquier herramienta para su desarrollo, lo que implica que el valor fundamental está en haber concebido dicho esquema conceptual.

En relación a la implementación, es importante indicar que un punto clave en el proceso de desarrollo e implementación fue el análisis de los datos proporcionados por el estudio de PISA, lo que permitió establecer que HISEI y HISCED serían los indicadores implementados. Cabe destacar que en esta etapa también se descartaron otros posibles indicadores por no tener la continuidad necesaria de datos dentro de todas las versiones del estudio PISA lo que significa no poder mostrar tendencias en el tiempo. Si bien, los resultados se centraron en el análisis de la información de Chile, el esquema conceptual del Data Warehouse PISA es aplicable a cualquier país del estudio. Por lo cual el diseño es reutilizable.

Luego, si buscamos un nuevo caso de aplicación de Data Warehouse con datos de estudios educacionales y/o pruebas estandarizadas podríamos utilizar la información del Sistema de Medición de la Calidad de la Educación (SIMCE) en Chile para encontrar las tendencias históricas y luego comparar los resultados. Examinando cómo las acciones o políticas públicas en educación tienen relación con las tendencias encontradas.

También encontramos proyección de esta investigación en la prueba estandarizada TIMSS, en este caso los tipos de prueba son Matemática y Ciencia. De este modo, se proyectan claramente nuevos casos de uso del esquema conceptual concebido para datos de pruebas estandarizadas por lo cual podríamos indicar que es un modelo dada su generalidad de uso en el ámbito educacional.

De esta forma, se ha mostrado que es posible utilizar información de estudios de gran escala para ser analizada con la tecnología que provee un Data Warehouse con el objetivo de realizar seguimientos y revisiones a las tendencias históricas. Por eso las proyecciones de este trabajo son variadas, ya que, por ejemplo, a nivel nacional en Chile se podría utilizar el mismo esquema conceptual del Data Warehouse propuesto con los datos del estudio SIMCE. Y a nivel internacional con las pruebas TIMSS y PIRLS.

## 6. Conclusiones

Se ha presentado un análisis multidimensional, de carácter preliminar, de los datos de la prueba estandarizada PISA usando un Data Warehouse. Este análisis debió potenciar competencias tecnológicas con competencias de análisis en el ámbito

educacional, social y de políticas públicas. La unión de estas competencias nos permite concluir que:

- Respecto al análisis de los datos de una prueba estandarizada como lo es PISA se indica:
  - Se ha propuesto un esquema conceptual para representar el Data Warehouse de la prueba estandarizada PISA. Dicho esquema conceptual puede ser reutilizado como modelo en otros estudios, según se señaló en la discusión. La reutilización depende de que las dimensiones de interés sean estudiantes, tipo de prueba, situación socioeconómica, escolaridad de padres y tiempo. En consecuencia, se puede hablar de un modelo para analizar información de pruebas estandarizadas a través del uso del Data Warehouse como herramienta de apoyo al seguimiento de las políticas públicas.
- Respecto a los resultados:
  - Permiten representar tendencias históricas a través de las dimensiones. Lo anterior requirió un proceso de análisis previo riguroso, respecto de la información que entrega la OECD a través de los resultados de la prueba PISA. Este análisis riguroso permitió generar el modelo que se indicó en el punto anterior.
- Respecto del uso de la tecnología que proveen los Data Warehouse (también llamados Data Warehouse) para analizar la información de estudios de gran escala como lo es la prueba estandarizada PISA se indica:
  - El mayor potencial está dado por la rapidez que otorga la tecnología para procesar una gran cantidad de datos. Es decir, una de las ventajas principales de este tipo de tecnología se basa en la estructura de la información que es multidimensional. Esta estructura permite realizar consultas relacionadas con indicadores y el tratamiento jerarquizado. Los reportes que se generan a través del Data Warehouse permiten apoyar a la toma de decisiones que se basan en un análisis de naturaleza multidimensional.

En relación a las preguntas de investigación se concluye que:

- Fue posible generar un esquema conceptual que sirve de modelo para interpretar los resultados históricos de pruebas estandarizadas usando como herramienta los Data Warehouse.
- También se mostró fue posible relacionar los resultados de un análisis multidimensional de los datos de PISA con las políticas públicas en el área educación.
- Es posible utilizar la información de PISA para encontrar tendencias históricas de los estudiantes respecto de la variable socioeconómica y estudio de los padres.

Como trabajo futuro se cargarán los datos de PISA del año 2012 en el Data Warehouse, lo que no implica realizar cambios al modelo conceptual, sino que sólo nos permitirá continuar analizando las tendencias históricas. Pues como se explicó en el desarrollo de esta investigación el esfuerzo está en generar el esquema conceptual que sirve como modelo para las investigaciones proyectadas.

## 7. Referencias

- [1] Baker, R., Yacef, K. (2009) The State of Educational Data Mining in 2009: A Review and Future Visions, *Journal of Educational Data Mining*, 1(1), 3–17
- [2] Berson, A., Smith, S. J. (1997) *Data Warehousing, Data Mining, and Olap*, McGraw-Hill
- [3] Calvet, L., Juan, Á. (2015) Educational Data Mining and Learning Analytics: differences, similarities, and time evolution, *RUSC. Universities and Knowledge Society Journal*, 12(3), 98-112
- [4] Chaudhuri, S., Sand, D., Umeshwar (1997) An overview of data warehousing and OLAP technology, 26(1) 65–74
- [5] Cravero, A., Sepúlveda, S. (2012) A chronological study of paradigms for datawarehouse design, *Ingeniería e Investigación*, 32(2), 58-62
- [6] De Garay, A., & del Valle-Díaz-Muñoz, G. (2012). Una mirada a la presencia de las mujeres en la educación superior en México. *Revista iberoamericana de educación superior*, 3(6), 3-30.
- [7] Donoso, S., Donoso, G. (2009) Políticas de gestión de la educación pública escolar en Chile (1990 -2010): una evaluación inicial Ensaio: aval. pol. públ. Educ., Rio de Janeiro, 17(64), 421-448.
- [8] Donoso, S., (2005) Reforma y política educacional en Chile 1990-2004: El neoliberalismo en crisis. *Estudios Pedagógicos* 31(1), 113-135.
- [9] Farias, C. (2012). Estrategias de aprendizajes en alumnos de primero medio en tres colegios de Talca y su relación con los rendimientos académicos. *Convergencia Educativa*, 1, 35-53. Recuperado de <http://www.convergenciaeducativa.cl/principal/wp-content/uploads/01-estrategias.pdf>
- [10] Fernández, O., Martínez-Conde, M., Melipillán, R. (2009). Estrategias de aprendizaje y autoestima. Su relación con la permanencia y deserción universitaria. *Revista Estudios Pedagógicos*, 35(1), 27-45. Recuperado de [http://www.scielo.cl/scielo.php?pid=S0718-07052009000100002&script=sci\\_arttext](http://www.scielo.cl/scielo.php?pid=S0718-07052009000100002&script=sci_arttext)
- [11] Galindo, A., García, H. (2010) Minería de datos en la Educación, Universidad Carlos III. Recuperado de: <http://www.it.uc3m.es/~jvillena/irc/practicas/10-11/08mem.pdf>
- [12] García, F., Fonseca, G., Concha, L. (2015). Aprendizaje y rendimiento académico en educación superior: Un estudio comparado. *Revista Actualidades Investigativas en Educación*, 15 (3), 1-26.
- [13] González Ramos, A. M., & Bosch, N. V. (2013). International mobility of women in science and technology careers: shaping plans for personal and professional purposes. *Gender, Place & Culture*, 20(5), 613-629.
- [14] Harinath, S., Zare, R., Meenakshisundaram, S., Carroll, M., & Lee, D. G. Y. (2011). Professional Microsoft SQL server analysis services 2008 with MDX. John Wiley & Sons.
- [15] Hernández, J., Ramírez, M. J., Ferri, C., (2004) *Introducción a la Minería de Datos*, Prentice Hall- Addison-Wesley
- [16] Herrera, F. (2014). Motivación, estrategias de aprendizaje y rendimiento académico, en alumnos de Nutrición y Dietética de la Universidad del Desarrollo. *Revista Educ Cienc Salud*, 11 (1), 38-46. Recuperado de <http://www2.udel.cl/ofem/recs/antiores/vol1112014/artinv11114e.pdf>
- [17] Inmon, W. H., Rudin, K., Buss, C. K., & Sousa, R. (1998). *Data warehouse performance*. John Wiley & Sons, Inc.
- [18] Kimball, R., Ross, M. (1998) *The Data Warehouse Toolkit*. 449
- [19] La Red, L., Martínez, M., Karanik, M., Giovannini, M., Pinto, N. (2015) Perfiles de Rendimiento Académico: Un Modelo Basado en Minería de Datos, *Revista Campus Virtuales*, 1(4)
- [20] Lopez Sancho, P., Alvarez-Marron, J., de Pablo, F., Masegosa Gallego, J., Mayoral Gaston, M. C., Molina Hernandez, E., & Sandalio Gonzalez, L. M. (2013). The women and science committee of the Spanish national research council: Ten years promoting gender equality and excellence *Arbor-Ciencia Pensamiento Y Cultura*, 189(759).
- [21] Luan, J. (2002) *Data Mining and Knowledge management in higher education. Potential applications*, Toronto
- [22] Malinowski, E., & Zimányi, E. (2009). Conceptual Modeling for Data Warehouse and OLAP Applications. *Encyclopedia of Data Warehousing and Mining* 293-300
- [23] Martínez, R., (2006) La metodología de los estudios PISA. *Revista de Educación, extraordinario*, 111-129
- [24] Martín, E., García, L., Torbay, A., Y Rodríguez, T. (2008). Estrategias de aprendizaje y rendimiento académico. *International Journal of Psychology and Psychological Therapy*, 8 (3), 401-412. Recuperado de <http://www.ijpsy.com/volumen8/num3/213/estrategias-de-aprendizaje-y-rendimiento-ES.pdf>
- [25] Mazón, J. N., & Trujillo, J. (2008). An MDA approach for the development of data warehouses. *Decision Support Systems*, 45(1), 41-58.
- [26] Pinninghoff, M.A., Salcedo, P., Contreras, R. (2007) *Neural Networks to Predict Schooling Failure/Success*, *Lecture Notes Computer Science*, 4528.
- [27] Pons-Peregort, O., Puig, M. D. C., Tura, M., & Illescas, C. M. (2013). Analysis of equal gender opportunity in science and technology: The professional careers of women scientists and technologists. *Intangible Capital*, 9(1), 65-90.
- [28] Reyes, Y., Nuñez, L. (2015) La inteligencia de negocio como apoyo a la toma de decisiones en el ámbito académico, *Revista Internacional de Gestión del Conocimiento y la Tecnología*, 3(2), 63-73.
- [29] Romero, C., Ventura, S., García, E. (2008) Data mining in course management systems: Moodle case study and tutorial, *Computers & Education*, 51(1), 368-384.
- [30] Romero, C., Ventura, S. (2011) *Data mining in education* Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(3).
- [31] Romero, O., Abello, A. (2009) A survey of Multidimensional Modeling Methodologies, *International Journal of Data Warehousing & Mining*, 5(2), 1-23.



- [32] Roux, R., Anzures, E. (2015). Estrategias de aprendizaje y su relación con el rendimiento académico en estudiantes de una escuela privada de educación media superior. *Revista Actualidades Investigativas en Educación*, 15 (1),1-16. Recuperado de <http://www.scielo.sa.cr/pdf/aie/v15n1/a14v15n1.pdf>
- [33] Salazar, J. M. y Leihy, P. S. (2013). El Manual Invisible: Tres décadas de políticas de educación superior en Chile (1980-2010). *Archivos Analíticos de Políticas Educativas*, 21(34). Recuperado de <http://epaa.asu.edu/ojs/article/view/1127>
- [34] Verger, A., Bonal X., y Zancajo. A. (2016). Recontextualización de políticas y (cuasi)mercados educativos. Un análisis de las dinámicas de demanda y oferta escolar en Chile. *Archivos Analíticos de Políticas Educativas*, 24(27), Recuperado de <http://dx.doi.org/10.14507/epaa.24.2098>
- [35] Winter, R., Strauch, B. (2003) A method for demand driven information requirements analysis in data warehousing projects, *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, 9-14
- [36] Zambrano, C., Rojas, D., Carvajal, K., Acuña, G. (2011) Análisis de rendimiento académico estudiantil usando data warehouse y redes neuronales, *Revista Chilena de Ingeniería*, 19(3),369–381
- [37] Zambrano, C., Varas, M., Urrutia, A. (2012) Enfoque MDA para el diseño de un Data Warehouse Difuso. *Ingeniare Revista Chilena de Ingeniería*, 20(1), 99-113
- [38] Zambrano, C., Rojas D., Varas M. (2011) "Data Warehouse con geolocalización y clustering". *Congreso Internacional de Informática Educativa*. Santiago, Chile.