

Un Método para Analizar Datos de Pruebas Educativas Estandarizadas usando Almacén de Datos y Triangulación

Carolina del C. Zambrano*, Darío F. Rojas y Pedro A. Salcedo

Departamento de Metodología de Investigación e Informática Educativa, Facultad de Educación
Universidad de Concepción, Concepción, Chile. (e-mail: carolinazambrano@gmail.com)

* Autor a quien debe ser dirigida la correspondencia

Recibido Sep. 15, 2017; Aceptado Nov. 24, 2017; Versión final Feb. 1, 2018, Publicado Ago. 2018

Resumen

Este trabajo, tiene como objetivo proponer un método para analizar datos de pruebas educativas estandarizadas usando Almacén de Datos (AD) y triangulación. Para lograr este objetivo se generó un esquema conceptual de AD que sirve de modelo para analizar los resultados históricos de pruebas estandarizadas. Se utiliza como caso de estudio la prueba PISA realizada por la Organización para la Cooperación y el Desarrollo Económico, OECD. Además, el resultado del análisis es relacionado por medio de la triangulación con las políticas públicas en el área educación en Chile. Desde un punto de vista educacional, los análisis presentados muestran como resultado el beneficio del análisis multidimensional y la triangulación que sirven para realizar un seguimiento histórico de los resultados obtenidos por diversos países en la prueba PISA. Todo lo anterior se logra estableciendo un método que permite replicar el proceso y aplicarlo en distintas pruebas educativas estandarizadas.

Palabras clave: data warehouse; pruebas estandarizadas; políticas públicas; triangulación; educación

A Method for Analyzing Data from Standardized Educational Tests using Data Warehouse and Triangulation

Abstract

This work proposes a method to analyze data from standardized educational tests using Data Warehouse (DW) and triangulation. To achieve this objective, a conceptual framework of DW was generated which serves as a model to analyze the historical results of standardized tests. The PISA test carried out by the OECD is used as a case study. In addition, the result of the analysis is related through the triangulation with public policies in education in Chile. From an educational point of view, the proposed analyzes show as a result the benefit of the multidimensional analysis and the triangulation that serve to carry out a historical follow-up of the results obtained by various countries in the PISA test. All the above is achieved by establishing a method that allows replicating the process and apply it in different standardized educational tests.

Keywords: data warehouse; standardized tests; public policy; triangulation; education

INTRODUCCIÓN

En la actualidad, las organizaciones educacionales poseen una gran cantidad información local proveniente de sus procesos académicos, procesos de gestión curricular, procesos de diagnóstico de estudiantes, entre otros. Así mismo, en el ámbito internacional existen pruebas estandarizadas como PISA (Programme for International Student Assessment) que representa una fuente importante de información educacional. PISA es realizada por la OECD (Organisation for Economic Cooperation and Development). La prueba PISA está dirigida a estudiantes de 15 años (Martínez, 2006) 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. El objetivo de PISA es comparar los resultados de aprendizaje de estudiantes en los distintos sistemas educativos de países miembros de la OECD. En Chile PISA se aplica desde el año 2000. En este contexto, otros ejemplos de pruebas estandarizadas internacionales son TIMSS (Trends in International Mathematics and Science Study) y PIRLS (Progress in International Reading Literacy Study).

Desde la perspectiva de pruebas estandarizadas nos preguntamos, ¿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 DW? ¿Estos resultados se pueden relacionar con las políticas públicas en el área educación del país Chile? ¿Es posible generar un método que permita triangular información educacional proveniente de estudios que analizan pruebas estandarizadas? ¿Es posible generar un esquema conceptual que sirva de modelo para interpretar los resultados históricos de pruebas estandarizadas usando los DW? Para dar respuesta a estas interrogantes se pueden utilizar herramientas de Minería de Datos Educacional (MDE) tal como DW (Chaudhuri y Umeshwar, 1997; Kimball y Ross, 2013) y/o minería de datos (Hernández et al, 2004), aunque cabe precisar que en esta investigación sólo nos centraremos en los DW. Así, podemos considerar que realizar análisis sobre los datos de los resultados de pruebas estandarizadas como PISA 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 usando un DW.

En consecuencia, dadas las ventajas de aplicar el enfoque de MDE ya sea mediante DW y/o minería de datos es que resulta interesante plantear este trabajo, de carácter preliminar, que tiene como objetivo aportar con un esquema conceptual que sirve como modelo para analizar los resultados históricos de pruebas estandarizadas en el marco de la propuesta de un método de análisis de datos que incluye tres etapas asociadas al diseño e implementación de un DW y como última etapa la triangulación de los resultados con estudios cualitativos. Se analizan dos indicadores incluidos en PISA que son HISEI y HISCED. HISEI corresponde al mayor índice socioeconómico de los padres y HISCED representa al mayor grado de educación alcanzado por los padres. Estos dos 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 DW 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 la prueba estandarizada PISA. Por ello, esta investigación puede abrir nuevas líneas de investigación que permitan analizar los resultados históricos de estudios que usan variables cognitivo-motivacionales, entre otros.

Data Warehouse

Según Chaudhuri (1997) un DW es una colección de datos orientada a un determinado ámbito, por ejemplo, al ámbito educacional. 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 DW son dimensiones, hechos y medidas (Chaudhuri y Umeshwar, 1997; Kimball y Ross 2013; Inmon et. al, 1998). Una dimensión permite agrupar información respecto a un área o un aspecto para el análisis. Algunos ejemplos de dimensiones en el ámbito educacional 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 y permite realizar el seguimiento en periodos de tiempo. Los hechos representan un cruce de información entre dos o más dimensiones, 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 situación socioeconómica”, en este caso la medida es cantidad de estudiantes y las dimensiones cruzadas son tipo de prueba, estudiante y situación socioeconómica. En este sentido, un DW 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 DW es cuantitativo.

Investigación en MDE

En el contexto de análisis de datos educativos mediante DW y/o minería de datos, cabe señalar que las instituciones de educación superior pueden utilizar la información que se acumula producto de sus procesos curriculares, académicos, entre otros, con el fin de mejorar la gestión educativa (Luan, 2002; Eckert y Suénaga, 2015). De este modo, nace el concepto de Minería de Datos Educativa (MDE) (Baker y Yacef, 2009; Romero y Ventura, 2011) que otorga un sinnúmero de ventajas y oportunidades para el ámbito educativo. No obstante, lo anterior, deseamos destacar que ningún autor ha estudiado la utilidad de los DW 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 lo anterior, podemos dar cuenta de investigaciones relacionadas con DW en contextos educativos tales como los trabajos de Reyes y Nuñez (2015), La Red et. al. (2015), Zambrano et. al. (2011) y Zambrano et. al. (2012). 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 DW encontramos el trabajo de Romero (2008) que aplica minería de datos sobre la información generada en el entorno virtual de aprendizaje Moodle y el trabajo de Pinninghoff et. al. (2007) que utiliza redes neuronales para predecir el éxito académico de estudiantes usando los datos de PISA. Adicionalmente, los trabajos de Mohamad y Tasir, (2013), Peña-Ayala (2014) y Shahiri e Husain (2015) han mostrado el estado de avance respecto de minería de datos educativa realizando revisiones de los artículos del área de EDM publicados en los últimos 10 años. No obstante, en estas revisiones no se evidencia investigaciones que propongan trabajar con los datos de estudios de pruebas estandarizadas en un DW, ni la creación de un método para dicho análisis, ni la incorporación de la triangulación con estudios cualitativos que permiten dar un sentido en el contexto particular a los resultados numéricos obtenidos por el análisis histórico de los datos usando el DW.

METODO PROPUESTO

El método propuesto se compone de 4 etapas que se observan en la Fig. 1. Es importante mencionar que para validar el método se utilizan los datos de PISA que son de libre disposición y que se pueden descargar desde la página web de la OECD. La muestra está dada por la cantidad de registros que se cargan durante la implementación del DW.

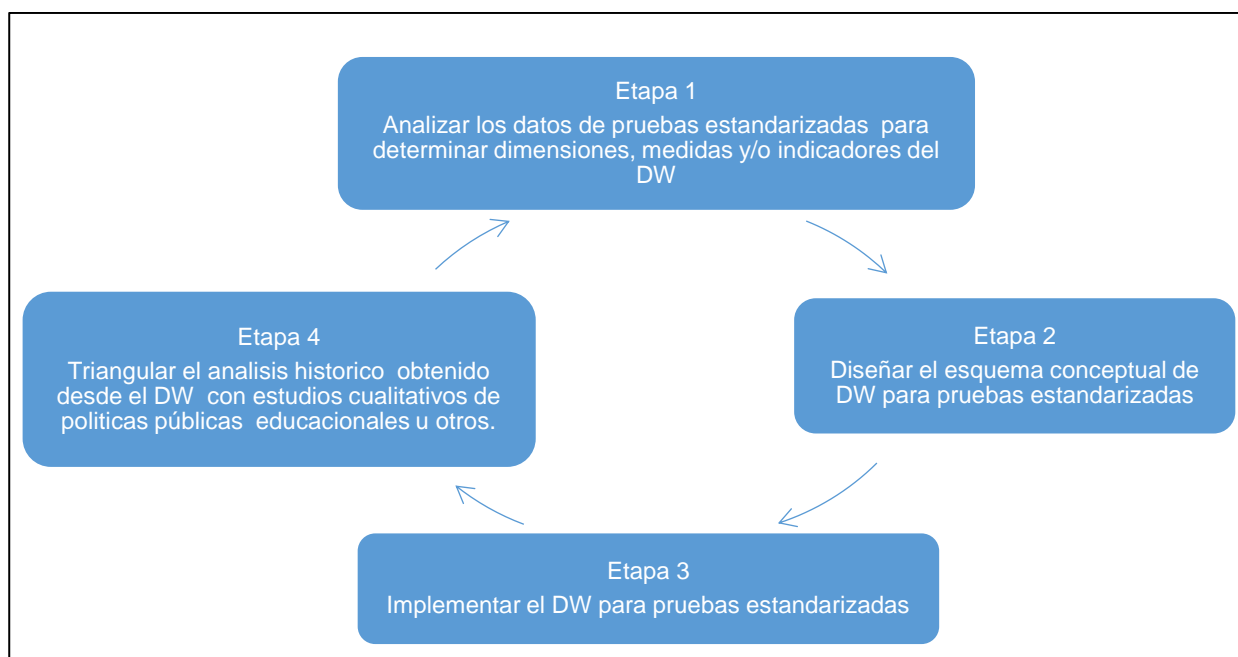


Fig. 1: Esquema del método propuesto, representa las etapas para analizar información desde pruebas estandarizadas usando un DW y triangulando con estudios cualitativos de políticas públicas educativas u otros.

Etapa 1: Analizar los datos de pruebas estandarizadas para determinar dimensiones y medidas

La primera etapa de la propuesta metodológica consiste en determinar dimensiones, medidas e indicadores desde los datos de pruebas estandarizadas. En esta investigación 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. En la Tabla 1 se muestra las dimensiones que se decidió analizar desde los datos fuente de PISA y en la Tabla 2 las medidas que se determinaron producto del riguroso análisis de las fuentes de PISA.

Tabla 1: Dimensiones del DW PISA

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

Tabla 2: Medidas del DW PISA

<i>Medidas e Indicadores</i>	<i>Descripción de las Medidas e Indicadores</i>
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 socioeconómico 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 DW. Y que, los índices HISEI y HISCED fueron seleccionados debido a que están presentes en todas las versiones (años 2000 a 2015) del estudio de PISA. Otros indicadores no estaban disponibles para poder realizar un análisis en el tiempo, pues no contaban con resultados en uno o más periodos de tiempo.

Etapa 2: Diseñar el esquema conceptual del DW de pruebas estandarizadas

En esta etapa, se diseña el esquema conceptual del DW de pruebas estandarizadas. Para realizar el diseño de un DW existen enfoques impulsados por los datos, enfoques impulsados por la demanda y enfoques híbridos (Cravero y Sepúlveda, 2012). En este caso se llevó a cabo un enfoque orientado a los datos, que consistió en analizar detalladamente las fuentes de datos de la prueba estandarizada PISA para determinar los elementos del DW tales como medidas y dimensiones que eran posibles de considerar con los datos disponibles. Es importante señalar que en el diseño de un DW se distinguen tres etapas que son diseño conceptual, diseño lógico y físico. Este último sirve para la implementación final antes de realizar el análisis multidimensional (Mazón y Trujillo, 2008; Malinowski y Zimányi, 2009). En esta investigación, el diseño conceptual del DW representa el modelo de DW para pruebas estandarizadas como PISA y es independiente de la implementación. Esta independencia ha sido estudiada y explicada en diversos trabajos de autores como Chaudhuri y Umeshwar (1997), Mazón y Trujillo (2008) e Malinowski y Zimányi (2009).

La notación que se utiliza para representar el esquema conceptual del DW para la prueba estandarizada PISA es la definida por Malinowski y Zimányi (2009) en el modelo MultiDim. La Fig. 2 muestra el esquema conceptual del DW para pruebas estandarizadas como PISA, en él se observan seis dimensiones, que representan las diferentes perspectivas desde las cuales se puede analizar la información de la prueba estandarizada PISA y cuatro indicadores que son HISEI, HISCED, cantidad de estudiantes y promedio en una prueba.

Etapa 3: Implementar DW para pruebas estandarizadas

Posterior al diseño, se implementa el DW para análisis de datos de pruebas estandarizadas. Esto implicó generar el esquema conceptual mostrado en la Fig. 2 a nivel lógico y físico en la herramienta Sql Server. En esta etapa se cargaron los registros históricos de la prueba estandarizada PISA para los años 2000, 2006, 2009, 2012 y 2015. Cabe señalar, que esta investigación busca aportar con un esquema conceptual genérico para pruebas estandarizadas, con un método y con el enfoque de triangulación.

El esquema lógico de la implementación se presenta en la Fig. 3. Para ello, se empleó el esquema en estrella con tal de simplificar las relaciones entre las dimensiones y la tabla de hecho, es decir, esta desnormalizado. Cabe mencionar que para implementar el DW se llevó a cabo un proceso ETL que consideró cuatro etapas que se muestran en la Fig. 4. La última tarea de la etapa 3 (proceso de implementación del DW para pruebas estandarizadas) es la exploración de la información mediante las distintas operaciones ROLAP. Mediante ROLAP se buscaron tendencias de interés de acuerdo con los objetivos planteados en esta investigación. La aplicación de ROLAP es consecuente con el diseño lógico en estrella que se escogió y se mostró en la Fig. 3. La herramienta que se utilizó para implementar el DW es SQL Server (Harinath et. al, 2011). Los resultados del análisis ROLAP se exportaron a Excel y se generó los gráficos que se muestran en la siguiente etapa 4.

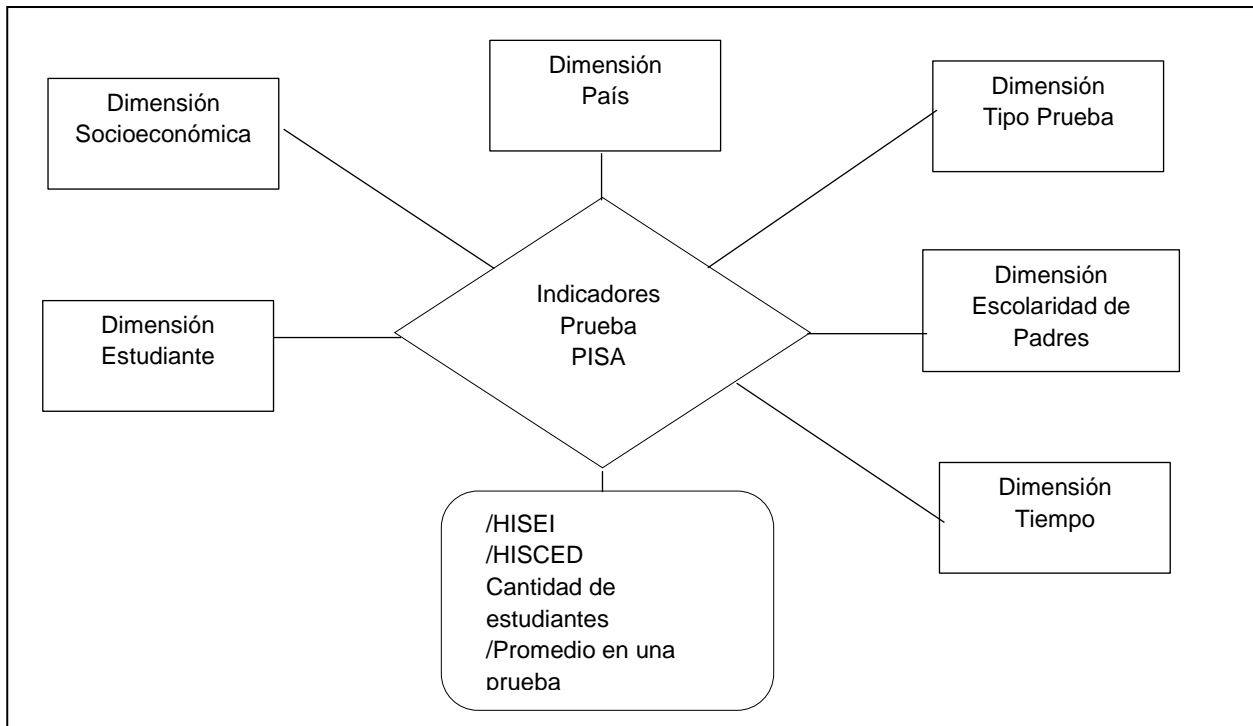


Fig. 2: Esquema Conceptual de DW para analizar prueba estandarizada, usando notación de modelo MultiDim (Malinowski y Zimányi, 2009).

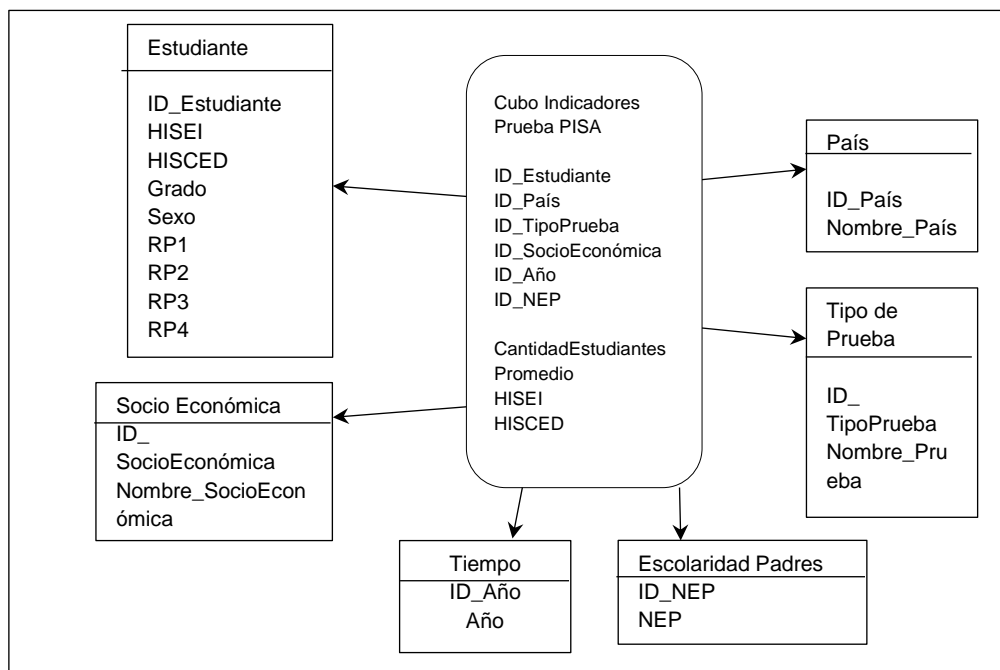


Fig. 3: Esquema Lógico, en estrella, de la implementación

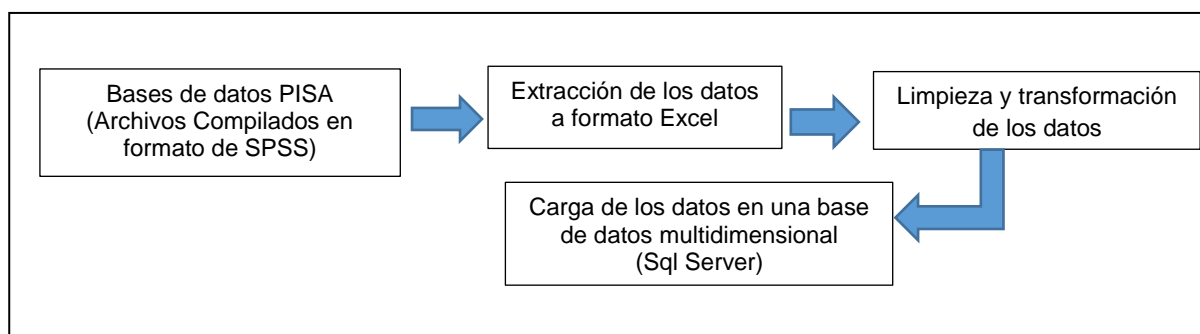


Fig. 4: Etapas del proceso ETL

Etapa 4: Triangular análisis obtenido desde el Data Warehouse con estudios cualitativos u otros

Una vez implementado el DW, 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 DW. Es importante indicar que la dimensión tiempo siempre estará presente en un análisis multidimensional usando DW, ya que ésta es la que permite representar las tendencias en el tiempo. De este modo, la etapa 4 presenta la proyección del análisis multidimensional en las políticas públicas. Esto se logra revisando estudios de carácter cualitativo para triangular (Bericat, 1998) la información cuantitativa que otorga el DW a través de las tendencias históricas con los datos cualitativos de estudios de políticas públicas en educación en Chile. Para realizar la triangulación se utilizan principalmente los siguientes estudios cualitativos relacionados con el análisis de políticas públicas en educación en Chile (Donoso, 2005; Donoso y Donoso, 2009; Cox, 2012; Salazar y Leihy, 2013; Verger et al., 2016).

Como estudio de caso se muestra el país Chile para proyectar el análisis realizando la triangulación. En efecto, se muestran reportes generados a través de la implementación del DW para pruebas estandarizadas. Así, la Fig. 5 muestra los puntajes promedio agrupados por nivel socioeconómico en Chile para los años 2000, 2006, 2009, 2012 y 2015. Se observa que existe una diferencia de más de 100 puntos entre los niveles socioeconómicos más bajos y los más altos. Esto nos sugiere que el nivel socioeconómico es un factor que incide en los resultados académicos de los estudiantes como se ha mostrado en estudios tales como (Pinninghoff et al., 2007; Verger et al., 2016) entre otros. Adicionalmente, se puede señalar que estos resultados corroboran lo expuesto en Bravo et. al (2015) respecto a que las expectativas de los padres y el tipo de escuela son factores que inciden en el rendimiento académico de los estudiantes. En efecto, en Chile el sistema educacional está segmentado en establecimientos municipales, establecimientos subvencionados y establecimientos particulares, y los estudiantes asisten a tipo de establecimiento dependiendo de la capacidad de pago de sus padres.

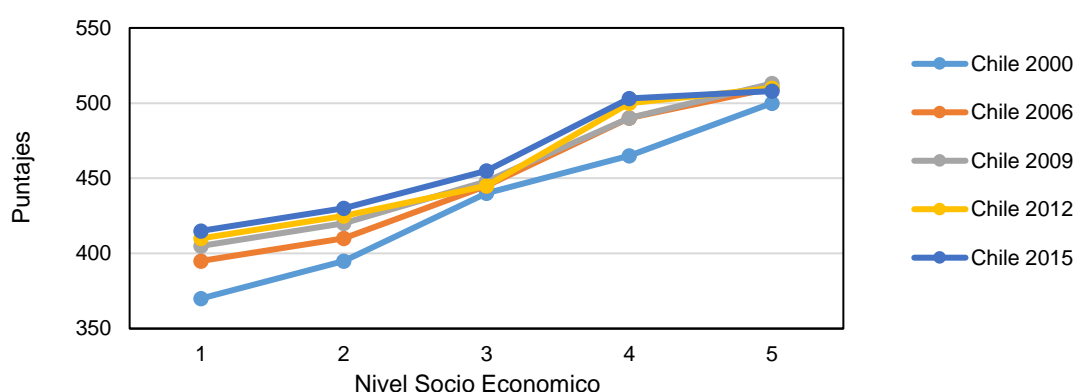


Fig. 5: Puntajes Promedio por Nivel Socioeconómico

Ahora bien, ¿qué sucede con los puntajes al relacionarlos con el nivel de escolaridad de los padres y el país Chile? La respuesta a esta pregunta se observa en la Fig. 6 donde se muestran los puntajes promedio agrupados por el nivel de escolaridad alcanzado por los padres de los estudiantes en Chile. Los niveles están representados desde 0 a 6 asociados a las siguientes etiquetas: 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; y Nivel 6: Postgrado. En la Fig. 6, se observa que en Chile existe relación entre el puntaje obtenido por el estudiante y el nivel de escolaridad de los padres lo que significa que a mayor nivel de escolaridad de los padres los puntajes de sus hijos aumentan. En efecto, si se asocian los resultados mostrados en las Fig.

5 y Fig. 6, se obtiene el mismo patrón en la tendencia histórica, que refleja la agrupación que la sociedad chilena ha dado su sistema educacional, donde a mayores recursos económicos y a mayor nivel de escolaridad de los padres, se han evidenciado mejores resultados de sus hijos en los puntajes obtenidos en pruebas estandarizadas como PISA. Esta situación se advierte en el estudio de políticas públicas en educación en Chile como el de Salazar y Leihy (2013), y también Cox (2012) sugiere algunas consecuencias de esta situación.

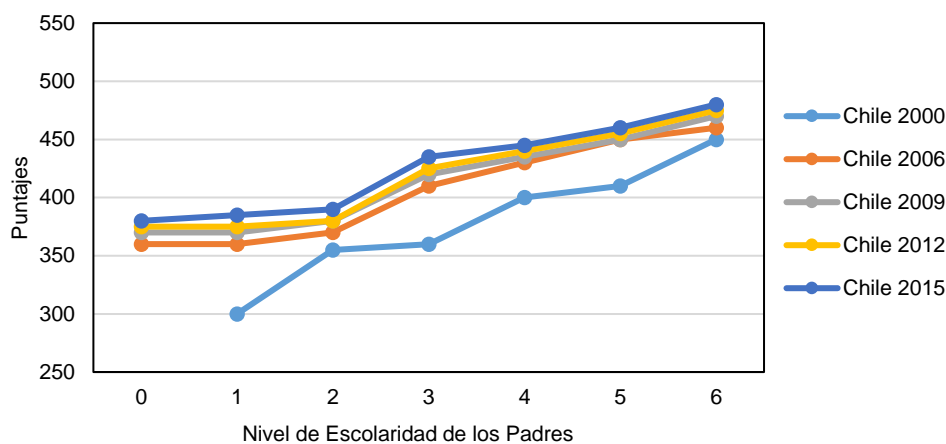


Fig. 6: Puntajes Promedios por Nivel de Escolaridad de los Padres

La Fig. 7 muestra la evolución de los puntajes promedio de las distintas pruebas para los años 2000, 2006, 2009, 2012 y 2015. Así, se observa que el resultado en la prueba de Lenguaje presenta el mayor progreso con una diferencia de 50 puntos entre el año 2000 y el 2009. Sin embargo, en 2012 el puntaje promedio en Lenguaje fue de 441 lo que marco un estancamiento que posteriormente se revierte en 2015 al subir a 459. En el área de Ciencias existe una mejora de 30 puntos entre el año 2000 y el 2009 que luego se estanca en las mediciones de los años 2012 y 2015 manteniéndose el puntaje en Ciencias en 447 al año 2015. En el caso de Matemática se evidencian los resultados más bajos, partiendo de 384 en el año 2000 hasta llegar a 421 en año 2009. Este puntaje se ha mantenido en los años 2012 y 2015 con 423 en ambos casos. Algunas respuestas a esta situación de estancamiento, sobre todo en matemática y ciencias, se encuentran en el modelo educacional chileno altamente segmentado y en las políticas públicas de educación como lo señalan los siguientes estudios (Verger et al., 2016; Donoso, 2005; Donoso y Donoso, 2009, entre otros). En particular, en Chile existen hallazgos que indican que el grupo social (alto, medio y bajo) es equiparable al tipo de dependencia administrativa de los establecimientos (Municipal, Particular Subvencionado, Particular) en el sistema educativo chileno que es segmentado socialmente donde la escolaridad de los padres juega un rol importante (García-Huidobro y Bellei, 2003).

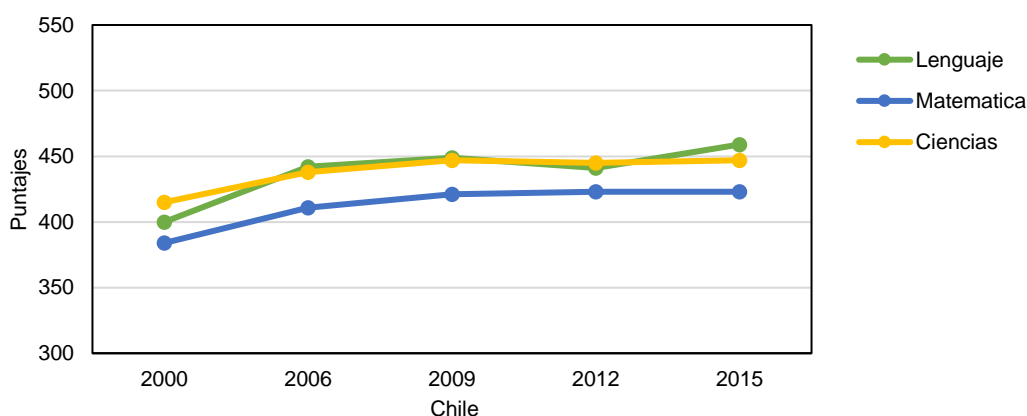


Fig. 7: Puntajes Promedio por Prueba para Chile

En la Fig. 8 se observa la evolución de los puntajes promedio para la prueba de Lenguaje por género entre los años 2000 y 2015 en Chile. Las mujeres mostraron una clara tendencia en alza en esta prueba, mientras que los hombres mantienen resultados más bajos. El año 2015 existe una diferencia significativa en el caso de las mujeres cuyo puntaje promedio llegó a 465. En el contexto de ciencias, la Fig. 9 muestra la tendencia de los puntajes promedio para esta prueba por género entre los años 2000 y 2015 en Chile. Tanto mujeres como hombres mostraron una tendencia positiva. No obstante, los hombres en promedio obtuvieron mejores resultados en esta prueba. Esta misma tendencia se observa en la Fig. 10 que presenta los resultados de la prueba de matemática por género.

Luego, en la Fig. 10 se presenta la evolución de los puntajes promedio para la prueba de matemáticas por género entre los años 2000 y 2015 en Chile. En esta prueba el puntaje de los hombres presentó una clara tendencia positiva, mientras que en el caso de las mujeres no fue así. La tendencia positiva en los resultados de los hombres respecto a los puntajes en ciencias y matemáticas es un claro reflejo de la baja cantidad de mujeres que estudian carreras del área ciencias y matemática. En este sentido, estudios relacionados con la presencia de mujeres en educación superior coinciden con esta tendencia de baja relación entre ciencia, matemática y mujeres. Algunas investigaciones que incluyen causas de la baja presencia de mujeres en ciencia y tecnología son De Garay y del Valle-Díaz-Muñoz (2012) y Pons-Peregort et. al (2013). Por ello, una tarea importante y un trabajo aún pendiente es fomentar la presencia de mujeres en ciencias, matemáticas e ingeniería como lo establecen González et. al. (2013) y López Sancho et. al. (2013).

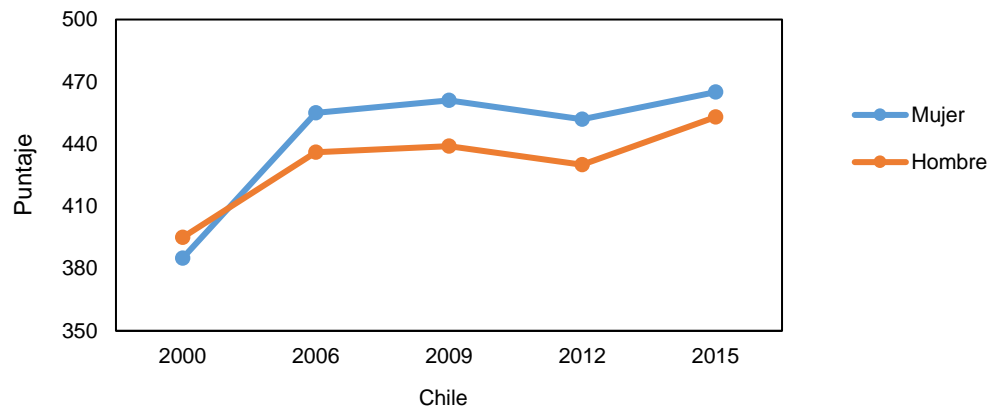


Fig. 8: Puntajes Promedios de Lenguaje por Género

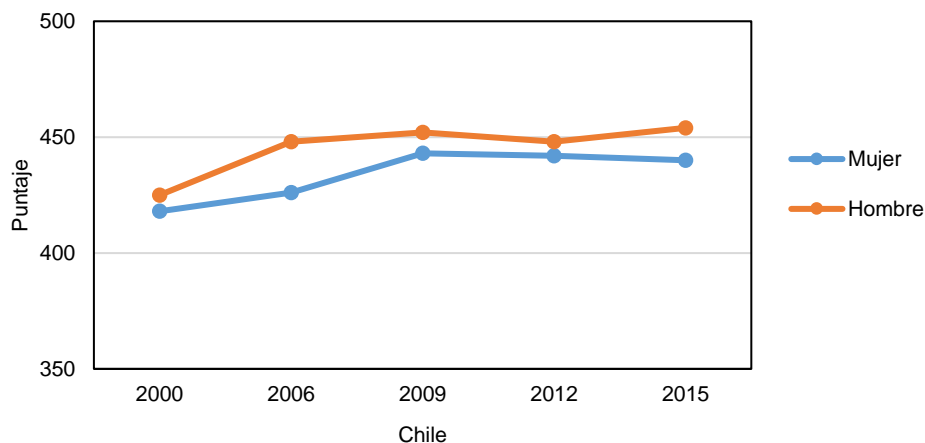


Fig. 9: Puntajes Promedio en Ciencias por Género

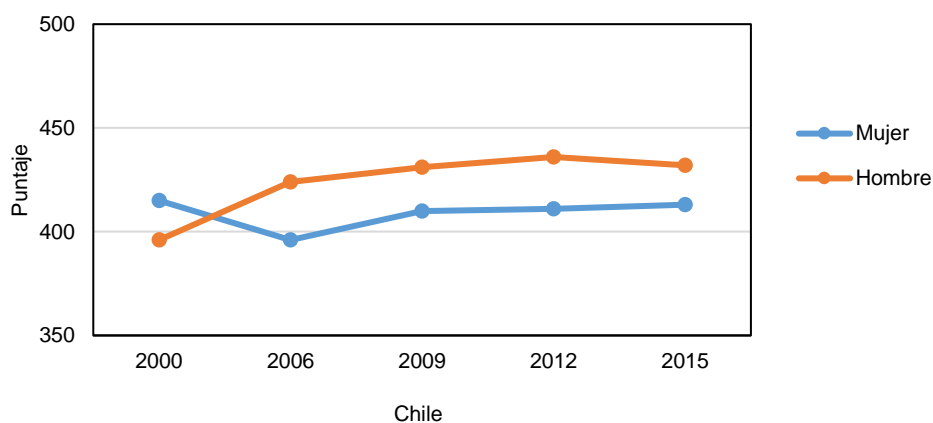


Fig. 10: Puntajes Promedios en Matemáticas por Género

En suma, los resultados nos proporcionan un primer acercamiento a la potencialidad de establecer una triangulación entre los resultados cuantitativos obtenidos a través de un DW y estudios de carácter cualitativo. Por ello, podemos mencionar que el enfoque propuesto a través del método permite unir dos tipos de estudios generando un enfoque de investigación mixto (Creswell y Plano Clark, 2007). Ya que usando la tecnología de análisis de datos que provee el DW y OLAP se tratan los datos cuantitativos que luego se explican por el análisis de estudios cualitativos. De esta forma, se aporta con una nueva idea sobre cómo aplicar procesos y métodos en el área de minería de datos educacional.

Desde la perspectiva del DW 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 tal como PISA, TIMSS, PIRLS y SIMCE. Cabe señalar que, la tecnología que permite procesar la información del DW implementado es OLAP (On-Line Analytical Processing). La tecnología OLAP es la que permitirá potencialmente soportar la gran cantidad de datos históricos que se cargarán en el tiempo en el DW diseñado. Es decir, cuando se agreguen los datos del estudio PISA 2018 no será necesario redefinir el esquema conceptual ni el método propuesto pues éstos ya han sido concebidos como un modelo para analizar los datos de pruebas estandarizadas. Metodológicamente esta independencia entre el diseño conceptual y la implementación representan una gran ventaja (Mazón y Trujillo, 2008). Desde el punto de vista de gestión, la triangulación entre análisis cuantitativos y cualitativos ayuda a crear y realizar seguimientos con énfasis en la calidad. Lo anterior, es producto de la intersección de resultados entre ambos tipos de análisis. Así, los resultados cuantitativos que permite obtener un DW pueden ser explicados a través de estudios cualitativos que le dan sentido a los patrones y tendencias numéricas que se obtienen producto del análisis multidimensional en el tiempo.

DISCUSION

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. Adicionalmente, mediante el método propuesto los resultados se pueden relacionar con las políticas públicas en el área educación del país en estudio que fue Chile. La relación se logra mediante la triangulación entre resultados cuantitativos obtenidos del DW y estudios cualitativos que explican las políticas públicas en educación durante los periodos analizados de la prueba PISA. De este modo, las acciones que se llevan a cabo a través de las políticas públicas en educación de cada país se ven reflejadas de forma positiva o no en los resultados de las mediciones estandarizadas como PISA. En suma, se mostró que la información obtenida a través del análisis multidimensional se puede triangular (Bericat, 1998) con estudios de carácter cualitativo u otro respecto de las políticas públicas en el ámbito educacional. Esta opción representa nuevo enfoque de triangulación de información en los estudios educacionales y abre un sin número de nuevas posibilidades. Lo anterior, en el marco del método propuesto que organiza el proceso de análisis y que cobra importancia en el contexto de la gran cantidad de información que hoy está disponible para analizar (Rojas et al., 2017)

También se puede señalar que al analizar los resultados de Chile obtenidos a través del DW de PISA existe una clara relación de estos con las políticas públicas que dirigen el sentido de la gestión de la enseñanza obligatoria en Chile entre los periodos de 1990 y 2015. Por ello, se observa que existe coincidencia con lo que señala Donoso (2005) y Donoso y Donoso (2009) 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 ha ido mejorando su posición. Respecto de la utilidad de los DW en el ámbito de la educación coincidimos con otros autores como Aljawarneh (2016), Reyes y Nuñez (2015), La Red et. al (2015), Zambrano et. al (2011), Zambrano et. al (2017) respecto a sus ventajas como herramienta para realizar seguimientos en base a la información histórica que se almacena y procesa en el DW. En este contexto, el seguimiento que se obtiene como resultado del análisis multidimensional debe apuntar a tomar acciones en pro de mejorar algún aspecto en el ámbito educacional que se puede explicar desde la triangulación como se mencionó anteriormente.

En relación con 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 (2009). Por esta razón, al ser un esquema conceptual, se hace independiente de su implementación, es decir, se podría utilizar cualquier tecnología de DW para su desarrollo, lo que implica que un valor fundamental en esta investigación es haber concebido dicho esquema conceptual. Si bien, los resultados se centraron en el análisis de la información de Chile, el esquema conceptual del DW PISA es aplicable a cualquier país del estudio. Por lo cual el diseño es reutilizable y se puede extender a otras dimensiones de análisis que incluye PISA.

De esta forma, se ha mostrado que es posible utilizar información de estudios que provienen de pruebas estandarizadas para ser analizada con la tecnología que provee un DW 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 DW y metodología propuesta con los datos de la prueba estandarizada SIMCE. Y a nivel internacional con las pruebas TIMSS y PIRLS. Otra línea que se puede explorar es cómo incorporar dimensiones con información de variables asociadas a psicología tal como autoeficacia, autorregulación del aprendizaje (Zambrano, 2016), entre otras.

En síntesis, esta investigación corrobora la información que ha sido proporcionada por otros estudios como el análisis de algunos resultados de SIMCE (Meckes y Carrasco, 2010; Bravo et. al, 2015) que también han revelado las grandes inequidades del país en los resultados de aprendizaje y su relación con las disparidades y segmentación social de Chile. También se aporta con un esquema conceptual que sirve como modelo para analizar los resultados históricos de pruebas estandarizadas en el marco de la propuesta de un método de análisis de datos que incluye tres etapas asociadas al diseño e implementación de un DW y una última etapa relacionada con la triangulación de los resultados con estudios cualitativos.

CONCLUSIONES

Sobre la base de lo expuesto en esta investigación se ha presentado un esquema conceptual genérico de DW para pruebas estandarizadas, un método de análisis de datos que incluye tres etapas asociadas al diseño e implementación de un DW y una última etapa relacionada con la triangulación de los resultados con estudios cualitativos. De este modo, se concluye que:

- 1) La propuesta de esquema conceptual para representar el DW de la prueba estandarizada PISA puede ser reutilizada como modelo en otros estudios que utilicen datos de este tipo. En suma, se sugiere un modelo para analizar información de pruebas estandarizadas a través del uso del DW como herramienta de apoyo al seguimiento de las políticas públicas.
- 2) Respecto del uso de la tecnología que proveen los DW para analizar la información de estudios de gran escala como lo es la prueba estandarizada PISA se concluye que el mayor potencial está dado por la rapidez que otorga la tecnología para procesar una gran cantidad de datos.
- 3) En relación con las preguntas de investigación se concluye que, es posible generar un esquema conceptual que sirve de modelo para interpretar los resultados históricos de pruebas estandarizadas usando como herramienta los DW. También, que es posible relacionar los resultados de un análisis multidimensional de los datos de PISA con las políticas públicas en el área educación de Chile a través de la triangulación y generar un método de análisis de los datos de pruebas estandarizadas que es mixto.

REFERENCIAS

- Aljawarneh, I.M., Design of a data warehouse model for decision support at higher education: A case study, doi:10.1177/0266666915621105, Information Development, 32(5), 1691-1706 (2016)
- Baker, R. y K. Yacef, The State of Educational Data Mining in 2009: A Review and Future Visions, Journal of Educational Data Mining, 1(1), 3-17 (2009)
- Bericat, E., La integración de los métodos cuantitativo y cualitativo en la investigación social: Significado y medida, 1ª Ed., 100-120, Ariel, España (1998)
- Bravo, M., S. Salvo y C. Muñoz, Profiles of Chilean students according to academic performance in mathematics: An exploratory study using classification trees and random forests, doi: 10.1016/j.stueduc.2015.01.002, In Studies in Educational Evaluation, 44, 50-59 (2015)
- Chaudhuri, S. y D. Umeshwar, An overview of data warehousing and OLAP technology, doi:10.1145/248603.248616, ACM SIGMOD Record, 26(1), 65-74 (1997)
- Cox, C., Política y políticas educativas en Chile 1990-2010, Revista Uruguaya de Ciencia Política, 21(1), 13-43 (2012)
- Cravero, A. y S. Sepúlveda, A chronological study of paradigms for datawarehouse design, Ingeniería e Investigación, 32(2), 58-62 (2012)
- Creswell, J. W. y V. Plano Clark, Designing and conducting mixed methods research, 3ª Ed., 210-230, SAGE, EE.UU (2017)
- De Garay, A. y G. Del Valle-Díaz-Muñoz, Una mirada a la presencia de las mujeres en la educación superior en México, Revista iberoamericana de educación superior, 3(6), 3-30 (2012)
- Donoso, S. y G. Donoso, 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 (2009)

- Donoso, S., Reforma y política educacional en Chile 1990-2004: El neoliberalismo en crisis, doi:10.4067/S0718-07052005000100007, Estudios Pedagógicos, 31(1), 113-135 (2005)
- Eckert, K.B. y R. Suénaga, Análisis de Deserción-Permanencia de Estudiantes Universitarios Utilizando Técnica de Clasificación en Minería de Datos, doi:10.4067/S0718-50062015000500002, Formación Universitaria, 8(5), 03-12 (2015)
- García-Huidobro, J.E. y C. Bellei, Desigualdad educativa en Chile. 1ª Ed., 4-63, Universidad Alberto Hurtado, Santiago de Chile (2003)
- García, F., G. Fonseca y L. Concha, Aprendizaje y rendimiento académico en educación superior: Un estudio comparado, doi: 10.15517/aie.v15i3.21072, Revista Actualidades Investigativas en Educación, 15 (3), 1-26 (2015)
- González Ramos, A. M. y N. Bosch, International mobility of women in science and technology careers: shaping plans for personal and professional purposes, doi: 10.1080/0966369X.2012.701198, Gender, Place & Culture, 20(5), 613-629 (2013)
- Harinath, S., R. Zare, S. Meenakshisundaram, M. Carroll y D. Lee, Professional Microsoft SQL server analysis services 2008 with MDX, 1ª Ed., 117-160, Wiley Publishing, EE.UU (2011)
- Hernández, J., M.J. Ramírez y C. Ferri, Introducción a la Minería de Datos, 1ª Ed., 3-19, Pearson Educación, España (2004)
- Inmon, W. H., K. Rudin, C. Buss y R. Sousa, DW performance, 1ª Ed., 300-330, John Wiley & Sons, EE. UU (1998)
- Kimball, R. y M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3ª Ed., 60-80, John Wiley, EE.UU (2013)
- La Red, L., M. Martínez y otros tres autores, Perfiles de Rendimiento Académico: Un Modelo Basado en Minería de Datos, Revista Campus Virtuales, 1(4) (2015)
- López-Sancho, P., J. Alvarez-Marrón y otros siete autores, The women and science committee of the spanish national research council: Ten years promoting gender equality and excellence, doi:10.3989/arbor.2013.759n1011, Arbor-Ciencia Pensamiento y Cultura, 189(759) (2013)
- Luan, J., Data mining and its applications in higher education, doi: 10.1002/ir.35, New directions for institutional research, 2(113), 17-36 (2002)
- Malinowski, E. y E. Zimányi, Conceptual Modeling for DW and OLAP Applications, doi: 10.4018/978-1-60566-010-3, Encyclopedia of Data Warehousing and Mining, 293-300 (2009)
- Martínez, R., La metodología de los estudios PISA, Revista de Educación, Extraordinario, 111-129 (2006)
- Mazón, J. N. y J. Trujillo, An MDA approach for the development of data warehouses, doi: 10.1016/j.dss.2006.12.003, Decision Support Systems, 45(1), 41-58 (2008)
- Meckes, L. y R. Carrasco, Two decades of SIMCE: an overview of the National Assessment System in Chile, doi:10.1080/09695941003696214, Assessment in Education: Principles, Policy & Practice, 17(2), 233-248 (2010)
- Mohamad, S. K. y Z. Tasir, Educational data mining: A review, doi: 10.1016/j.sbspro.2013.10.240, Procedia-Social and Behavioral Sciences, 97, 320-324 (2013)
- Peña-Ayala, A. Educational data mining: A survey and a data mining-based analysis of recent works, doi: 10.1016/j.eswa.2013.08.042, Expert systems with applications, 41(4), 1432-1462 (2014)
- Pinninghoff, M.A., P. Salcedo y R. Contreras, Neural Networks to Predict Schooling Failure/Sucess, doi:10.1007/978-3-540-73055-2_59, Lecture Notes Computer Science, 4528 (2007)
- Pons-Peregort, O., M. Puig, M. Tura y C. Ilescas, Analysis of equal gender opportunity in science and technology: The professional careers of women scientists and technologists, doi:10.3926/ic.375, Intangible Capital, 9(1), 65-90 (2013)
- Reyes, Y. y L. Núñez, 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 (2015)
- Rojas, D., C. Zambrano y P. Salcedo, Metodología de Análisis de Disponibilidad Léxica en Alumnos de Pedagogía a través de la Comparación Jerárquica de Lexicones, doi:10.4067/S0718-50062017000400002, Formación Universitaria, 10(4), 03-14 (2017)
- Romero, C., S. Ventura y E. García, Data mining in course management systems: Moodle case study and tutorial, doi: 10.1016/j.compedu.2007.05.016, Computers & Education, 51(1), 368-384 (2008)
- Romero, C. y S. Ventura, Data mining in education, doi: 10.1002/widm.1075, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(3), 12-27 (2011)
- Salazar, J.M. y P. Leihy, El Manual Invisible: Tres décadas de políticas de educación superior en Chile (1980-2010), <https://epaa.asu.edu/ojs/article/viewFile/1127/1091>, ISSN: 1068-2341, Archivos Analíticos de Políticas Educativas, 21(34) (2013)
- Shahiri, A.M. y W. Husain, A review on predicting student's performance using data mining techniques, doi: 10.1016/j.procs.2015.12.157, Procedia Computer Science, 72, 414-422 (2015)
- Verger, A., X. Bonal y A. Zancajo, 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) (2016)

Winter, R. y B. Strauch, A method for demand driven information requirements analysis in data warehousing projects, doi:10.1109/HICSS.2003.1174602, Proceedings of the 36th Annual Hawaii International Conference on System Sciences, 9-14 (2003)

Zambrano, C., D. Rojas, K. Carvajal y G. Acuña, Análisis de rendimiento académico estudiantil usando DW y redes neuronales, doi:10.4067/S0718-33052011000300007, Revista Chilena de Ingeniería, 19(3), 369–381 (2011)

Zambrano, C., M. Varas y A. Urrutia, Enfoque MDA para el diseño de un DW Difuso, doi:10.4067/S0718-33052017000200242, Ingeniare Revista Chilena de Ingeniería, 20(1), 99-113 (2012)

Zambrano, C., Autoeficacia, Prácticas de Aprendizaje Autorregulado y Docencia para fomentar el Aprendizaje Autorregulado en un Curso de Ingeniería de Software, doi:10.4067/S0718-50062016000300007, Formación Universitaria, 9(3), 51-60 (2016)

Zambrano, C., A. Urrutia y M. Varas, Análisis de rendimiento académico estudiantil usando Data Warehouse Difuso, doi:10.4067/S0718-33052017000200242, Ingeniare. Revista chilena de Ingeniería, 25(2), 242-254 (2017)