

**UN MERCADO DE DATOS PARA EL ANÁLISIS MULTIDIMENSIONAL DE LAS
PRUEBAS SABER 5 DE LAS INSTITUCIONES EDUCATIVAS DE LOS
MUNICIPIOS DE LA SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE
NARIÑO**



Universidad de **Nariño**

**JONATHAN ALVEIRO BUCHELI CAICEDO
BAIRON YEFERSON IMBACUAN YANDUN**

**UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA DE SISTEMAS
IPIALES
2018**

**UN MERCADO DE DATOS PARA EL ANÁLISIS MULTIDIMENSIONAL DE LAS
PRUEBAS SABER 5 DE LAS INSTITUCIONES EDUCATIVAS DE LOS
MUNICIPIOS DE LA SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE
NARIÑO**

**JONATHAN ALVEIRO BUCHELI CAICEDO
BAIRON YEFERSON IMBACUAN YANDUN**

Trabajo de grado presentado como requisito parcial para optar el título de
Ingenieros de Sistemas

**PhD. SILVIO RICARDO TIMARAN PEREIRA
DIRECTOR**

**UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA DE SISTEMAS
IPIALES
2018**

NOTA DE RESPONSABILIDAD

“La universidad de Nariño no se hace responsable de las opiniones o resultados obtenidos en el presente trabajo y para su publicación priman las normas sobre el derecho de autor”.

Artículo 13, Acuerdo No. 005 del 26 de enero de 2010, emanado por el Honorable Consejo Académico de la Universidad de Nariño.

“Las ideas y conclusiones aportadas en el siguiente trabajo son responsabilidad exclusiva del autor.”

Artículo 1ro del Acuerdo No. 324 de octubre 11 de 1966 emanado del Honorable Consejo Directivo de la Universidad de Nariño.

NOTA DE ACEPTACIÓN

Firma del jurado

Firma del jurado

Firma del asesor

Firma del co-asesor

Ipiales, 15 de mayo de 2018

AGRADECIMIENTOS

A mis profesores quienes me transmitieron su conocimiento para poder ser un excelente profesional. Al profesor Ricardo Timaran que gracias a su experiencia, conocimiento y confianza permitió desempeñarme como investigador y sacar adelante nuestro proyecto. Y finalmente a la VIPRI quienes nos apoyaron con los recursos económicos para poder adquirir mayor conocimiento y fortalecer nuestro proyecto.

A la Universidad de Nariño quien fue mi segundo hogar, Al profesor Ricardo Timaran por permitirnos desarrollar nuestro proyecto y quien, con su conocimiento y su larga trayectoria en el campo investigativo, fomentó el espíritu investigativo en cada uno de nosotros. Y finalmente a los profesores quien nos prepararon y brindaron sus conocimientos para desempeñarnos en el campo laboral

“A mis padres, hermanos y sobrinas. A mis padres que siempre me han apoyado con sus consejos, confianza y comprensión, y son quienes me dan la fortaleza para mejorar día a día y seguir con cada uno de mis proyectos de vida. Mis hermanos por ser mi apoyo incondicional y brindarme todo su cariño y a mis sobrinas por llenar mi vida de alegría. Los amo.”

Jonathan Alveiro Bucheli Caicedo

“A mi madre, quien ha sido el eje de motivación y respaldo en mi carrera y me apoyo en la lucha de mis sueños, quien con sus consejos y oraciones fueron mi fortaleza para continuar y seguir adelante. Mis hermanas por ser confidentes de traspasos, desesperos y por brindarme su apoyo y amor. Las quiero mucho”

Bairon Yeferson Imbacuan Yandun

RESUMEN

En ese documento se presentan los resultados del proyecto de investigación que tuvo como objetivo construir un mercado de datos para almacenar la información histórica de los estudiantes de las Instituciones Educativas de los municipios de la subregión de Obando del departamento de Nariño que presentaron las pruebas saber 5° entre los años 2014 al 2016.

Se utilizó como metodología para diseñar el modelo multidimensional del mercado de datos, la propuesta de Kimball. El mercado de datos se integró a la herramienta de inteligencia de negocios libre Pentaho para realizar el análisis multidimensional y la visualización de resultados. Este mercado de datos brinda información limpia, transformada y de calidad que ayudara a los directivos de las Instituciones Educativas de los municipios de la subregión de Obando y a las Secretarías de Educación municipales, a la toma de decisiones en lo relacionado con la formulación de estrategias para el mejoramiento de la calidad educativas de las instituciones.

Palabras clave: Mercado de Datos, Análisis Multidimensional, Instituciones Educativas, Subregión de Obando.

ABSTRACT

In this document the results of research project aimed to build a data mart to store the historical information of the students of the Educational Institutions of the municipalities of the Obando subregion of the department of Nariño who presented the tests Saber 5th between the years 2014 to 2016, are presented.

As methodology was used the proposed by Kimball to design the multidimensional model of the data mart. The data mart was integrated into the Pentaho free business intelligence tool to perform multidimensional analysis and visualization of results. This data mart provides clean, transformed and quality information that will help the managers of the Educational Institutions of the municipalities of the Obando subregion and the Municipal Secretaries of Education, to make decisions regarding the formulation of strategies for the improvement of the educational quality of the institutions.

Keywords: Data Mart, Multidimensional Analysis, Educational Institutions, Obando Subregion.

CONTENIDO

	Pág.
INTRODUCCIÓN	16
1. FUNDAMENTOS TEÓRICOS	23
1.1. Data Warehouse	23
1.2. Datamart	25
1.3. Diseño de un Datamart	26
1.4. Análisis multidimensional	27
1.5. PostgreSQL	28
1.6. Pentaho.....	28
1.7. Modelo multidimensional en estrella	29
2. CONCEPTUALIZACIÓN DE PRUEBAS SABER 5°	31
2.1. Elementos conceptuales de las pruebas saber.....	31
2.1.1. Factores asociados:	34
2.1.1.1. Contexto.....	35
2.1.1.2. Insumos	39
2.1.1.3. Procesos	40
2.1.1.4. Resultados Educativos.....	42
2.1.2. Estructura y alineación de las pruebas saber.	43
3. METODOLOGÍA	47
3.1. Planificación.....	47
3.2. Análisis de requerimientos	47
3.3. Modelado dimensional	47
3.4. Diseño físico	48
3.5. Diseño del sistema de extracción, transformación y carga (ETL)	48
3.6. Especificación y desarrollo de aplicaciones de inteligencia de negocios bi (del inglés business intelligence)	48
4. IMPLEMENTACIÓN	49
4.1. Análisis de requerimientos.....	49
4.2. Modelado dimensional.....	49

4.3.	Diccionario de datos del Datamart.....	52
4.4.	Preparación de ambiente para el proceso ETL.....	54
4.5.	Diseño del sistema de extracción, transformación y carga (ETL)	57
4.5.1.	Fase de extracción.....	57
4.5.1.1.	Integración de Datos.....	57
4.5.1.2.	Análisis de la calidad de datos	61
4.5.1.3.	Limpieza de Datos.	71
4.5.2.	Fase de transformación.	71
4.5.2.1.	Transformación de Datos.....	71
4.5.2.2.	Eliminación de Datos.	72
4.5.3.	Fase de carga.	72
4.6.	Implementación de la bodega en Postgres	75
4.7.	Integración del Datamart con Pentaho BI	76
4.7.1.	Cubo de datos.....	76
4.7.2.	Operaciones OLAP	78
4.8.	Especificación y desarrollo de aplicaciones de BI.....	79
5.	RESULTADOS Y ANÁLISIS	82
	CONCLUSIONES Y RECOMENDACIONES	108
	REFERENCIAS	109
	ANEXOS.....	116

LISTA DE FIGURAS

	Pág.
Figura 1. Estructura lógica de una Bodega de Datos.....	24
Figura 2. Fases de la metodología Kimball.....	27
Figura 3. Esquema Tipo Estrella.....	30
Figura 4. Modelo CIPP.....	33
Figura 5. Marco de Factores Asociados.	34
Figura 6. Modelo Multidimensional tipo estrella.	51
Figura 7. Configuración del Path.....	55
Figura 8. Configuración de variables de entorno.	56
Figura 9. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión estudiante.	72
Figura 10. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión institución.....	73
Figura 11. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión municipio.....	73
Figura 12. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión prueba.....	74
Figura 13. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la tabla de hechos.	74
Figura 14. Implementación del Modelo Multidimensional tipo estrella en PostgreSQL.	75
Figura 15. Configuración del cubo para las consultas OLAP en Schema Workbench.....	77
Figura 16. Análisis multidimensional en BI server de Pentaho.	78
Figura 17. Pantalla de inicio Pentaho.	80
Figura 18. Creación de reporte en Pentaho.....	80
Figura 19. Sentencia SQL para diseño de cubos.....	81
Figura 20. No. estudiantes por municipio que presentaron las pruebas saber 5° entre 2014-2016.	82

Figura 21. Puntaje promedio de los resultados de matemáticas clasificados por Municipios Subregión de Obando.	83
Figura 22. Puntaje promedio de los resultados de Lenguaje clasificados por Municipios Subregión de Obando.	84
Figura 23. Puntaje promedio de los resultados de Ciencias clasificados por Municipios Subregión de Obando.	84
Figura 24. Puntaje promedio de los resultados de competencias clasificados por Municipios Subregión de Obando.	85
Figura 25. No. Estudiantes que presentaron las pruebas saber 5° clasificados por género.	86
Figura 26. Puntaje promedio de los resultados por municipio clasificados por género en el área de matemáticas.	87
Figura 27. Puntaje promedio de los resultados por municipio clasificados por género en el área de Ciencias.	88
Figura 28. Puntaje promedio de los resultados por municipio clasificados por género en el área de Ciencias.	88
Figura 29. Puntaje promedio de los resultados por municipio clasificados por género en el área de Competencias.	89
Figura 30. Puntaje promedio de la prueba de Matemáticas, clasificados por géneros.	90
Figura 31. Puntaje promedio de la prueba de Lenguaje clasificados por géneros.	92
Figura 32. Puntaje promedio de la prueba de ciencias, clasificados por géneros.	92
Figura 33. Puntaje promedio de la prueba de Competencias clasificados por géneros.	93
Figura 34. Puntaje promedio de los resultados clasificados por año área de matemáticas.	94
Figura 35. Puntaje promedio de los resultados clasificados por año área de Lenguaje.	94
Figura 36. Puntaje promedio de los resultados clasificados por año área de Ciencias.	95
Figura 37. Puntaje promedio de los resultados clasificados por año área de Competencias.	95
Figura 38. Puntaje promedio de los resultados de Matemáticas clasificados por tipo de establecimiento.	97
Figura 39. Puntaje promedio de los resultados de Lenguaje clasificados por tipo de establecimiento.	97

Figura 40. Puntaje promedio de los resultados de Ciencias clasificados por tipo de establecimiento.	98
Figura 41. Puntaje promedio de los resultados de Competencias clasificados por tipo de establecimiento.	98
Figura 42. Puntaje promedio de los resultados de Matemáticas clasificados por zona.	100
Figura 43. Puntaje promedio de los resultados de Lenguaje clasificados por zona.	100
Figura 44. Puntaje promedio de los resultados de Ciencias clasificados por zona.	101
Figura 45. Puntaje promedio de los resultados de Competencias clasificados por zona.	101
Figura 46. Puntaje de estudiantes con discapacidad cognitiva área de matemáticas clasificados por zona.	103
Figura 47. Puntaje de estudiantes con discapacidad cognitiva área de lenguaje clasificados por zona.	104
Figura 48. Puntaje de estudiantes con discapacidad cognitiva área de ciencias clasificados por zona.	105
Figura 49. Puntaje de estudiantes con discapacidad cognitiva área de competencias clasificados por zona.	106
Figura 50. Certificado del Encuentro nacional de estudiantes de ingeniería informática y afines.	129
Figura 51. Certificado del 3er congreso andino en computación, informática y educación.	130

LISTA DE TABLAS

	Pág.
Tabla 1. Resultados pruebas Saber 5° por áreas municipio de Ipiales.....	18
Tabla 2. Procesos de las competencias genéricas. Lenguaje Matemáticas Ciencias.	44
Tabla 3. Componentes de las competencias genéricas.....	45
Tabla 4. Dimensiones y granularidad del Datamart.	49
Tabla 5. Diccionario de Datos de la Dimensión Estudiantes.....	52
Tabla 6. Diccionario de Datos de la Dimensión Instituciones.	52
Tabla 7. Diccionario de Datos de la Dimensión Municipios.	53
Tabla 8. Diccionario de Datos de la Dimensión Prueba.....	53
Tabla 9. Diccionario de Datos de la Tabla de Hechos.	53
Tabla 10. Número de registros en cada repositorio.	57
Tabla 11. Valores Plausibles.....	58
Tabla 12. Entidades.	59
Tabla 13. Establecimientos.	60
Tabla 14. Municipios.	60
Tabla 15. Análisis de la calidad de Valores plausibles.....	61
Tabla 16. Análisis de la Calidad de Municipios.	65
Tabla 17. Análisis de la Calidad de Instituciones.	65
Tabla 18. Análisis de la Calidad de Dimensión prueba.	66
Tabla 19. Análisis de la Calidad de Dimensión Municipio.	67
Tabla 20. Análisis de la Calidad de Dimensión estudiante.	67
Tabla 21. Análisis de la Calidad de Dimensión institución.	68
Tabla 22. Análisis de la Calidad de Tabla de hechos.	70
Tabla 23. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las 4 áreas evaluadas por él ICFES.....	79
Tabla 24. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las áreas, clasificado por géneros.....	89
Tabla 25. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las 4 áreas, clasificado por periodo anual.....	96
Tabla 26. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las 4 áreas, clasificado por tipo de establecimiento.	99

Tabla 27. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las 4 áreas, clasificado por zona.....	102
Tabla 28. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las 4 áreas, clasificado por discapacidad cognitiva en las zonas geográficas.	107

INTRODUCCIÓN

La Ley 1324 le confiere al Instituto Colombiano para Evaluación de la Educación (ICFES) la misión de evaluar, mediante exámenes externos estandarizados, la formación que se ofrece en el servicio educativo en los distintos niveles. También establece que el Ministerio de Educación Nacional (MEN) define lo que debe evaluarse en estos exámenes [2].

Por su parte, en el Plan Decenal 2006-2016 se propuso “organizar, implementar y consolidar un sistema de seguimiento y evaluación del sector educativo, que dé cuenta de los logros y dificultades de los estudiantes, su acceso, cobertura y permanencia en el sistema y la eficiencia de los entes responsables de la prestación y la calidad del servicio” [5].

Para cumplir con lo anterior, el ICFES ha avanzado en la alineación del Sistema Nacional de Evaluación Externa Estandarizada (SNEE), a través de la reestructuración de los exámenes: en 2009 con un nuevo diseño de SABER 3°, 5° y 9°; en 2010 con el rediseño de SABER PRO; en 2014 con los cambios en SABER 11°. La alineación posibilita la comparación de los resultados en distintos niveles educativos, ya que los diferentes exámenes evalúan unas mismas competencias en algunas de las áreas que los conforman, a saber, las competencias genéricas [2].

El propósito de las pruebas SABER 3°, 5° y 9° es contribuir al mejoramiento de la calidad de la educación colombiana, mediante la realización de evaluaciones periódicas en las que se valoran las competencias básicas de los estudiantes y se analizan los factores que inciden en sus logros [3]. Las competencias en Lenguaje, Matemáticas, Ciencias Naturales y Competencias Ciudadanas son áreas, que a criterio de expertos se constituyen en los componentes principales para desarrollar en los estudiantes, las competencias básicas para la vida y que les permiten alcanzar un desempeño eficaz y eficiente dentro de la sociedad [1].

Los resultados de estas evaluaciones permitieron que los establecimientos educativos, las secretarías de educación, el Ministerio de Educación Nacional (MEN) y la sociedad en general conozcan cuáles son las fortalezas y debilidades

y, a partir de estas, puedan definir planes de mejoramiento en sus respectivos ámbitos de actuación.

Su carácter periódico posibilita, además, valorar cuáles han sido los avances en un determinado lapso y establecer el impacto de programas y acciones específicas de mejoramiento.

Uno de los principales problemas que presentaron las Instituciones Educativas de la Subregión de Obando, es el bajo rendimiento que se viene presentando en los últimos años en las pruebas SABER 5°. En la tabla 1 se presenta los resultados obtenidos en las pruebas SABER 5° por áreas en el año 2016 [4].

Tabla 1. Resultados pruebas Saber 5° por áreas municipio de Ipiales.

Áreas	Porcentaje insuficiente	Porcentaje mínimo	Porcentaje satisfactorio	Porcentaje avanzado
Lenguaje	9%	37%	37%	17%
Ciencias Naturales	6%	49%	29%	15%
Matemáticas	25%	35%	25%	15%
Promedio	13.33%	40.33%	30.33%	15.66%

Fuente: ICFES resultados por municipios año 2016.

Según los datos de la tabla 1, el 53.7 % de los estudiantes de los establecimientos educativos del municipio de Ipiales se encuentran en un nivel bajo de rendimiento en estas pruebas y el 46.34% se encuentran en porcentaje medio alto.

Para lograr que en diez años Colombia sea el país más educado de América Latina, se requiere que todos los colegios el personal docente, los estudiantes y los padres de familia sumen esfuerzos para que en una escala de 1 a 10, la primaria tenga un promedio de 7.01.

En este proyecto de investigación que tuvo como objetivo construir un mercado de datos que almacene la información histórica de los estudiantes de las Instituciones Educativas de la Subregión de Obando que presentaron las pruebas Saber 5° entre los años 2014 y 2016. Este mercado de datos brinda información limpia, transformada y de calidad que ayudará a los directivos de las Instituciones Educativas de la Subregión de Obando y de las Secretarías de Educación de estos Municipios, a la toma de decisiones en lo relacionado con la formulación de estrategias para el mejoramiento de la calidad educativa en estas instituciones.

PLANTEAMIENTO DEL PROBLEMA

Es función principal de la evaluación en la educación orientar y apoyar las acciones de mejoramiento de la calidad mediante la obtención, análisis e interpretación de información válida y confiable. Una de las dificultades presentes a lo largo de la historia del sistema educativo colombiano ha sido la falta de información válida y confiable requerida en el diseño y ejecución de políticas, planes, programas y proyectos, que en conjunto hacen posible hacer realidad los propósitos y esfuerzos del Gobierno y la sociedad para disponer de una educación de mejor calidad y equidad [1]. Por otra parte, los resultados de pruebas nacionales e internacionales muestran que Colombia posee un sistema educativo con bajos logros académico de sus estudiantes, en cada uno de los niveles de estudio [2].

Actualmente el ICFES diseña y aplica las pruebas Saber 3°, Saber 5°, Saber 9°, Saber 11°, con las cuales evalúa la Educación Básica y Media. Estas pruebas evalúan competencias básicas, es decir, competencias que se consideran indispensables sin importar la particularidad de las instituciones educativas. Su objetivo es contribuir al mejoramiento de la calidad de la educación colombiana, mediante la realización de evaluaciones periódicas (censales y muestrales) en las que se valoran las competencias básicas de los estudiantes y se analizan los factores que inciden en sus logros. Los resultados de estas evaluaciones permiten que los establecimientos educativos, las secretarías de educación, el Ministerio de Educación Nacional (MEN) y la sociedad en general conozcan cuáles son las fortalezas y debilidades y, a partir de estas, puedan definir planes de mejoramiento en sus respectivos ámbitos de actuación. Su carácter periódico posibilita, además, valorar cuáles han sido los avances en un determinado lapso y establecer el impacto de programas y acciones específicas de mejoramiento [3].

La Subregión de Obando o Sur es una de las 13 subregiones del departamento de Nariño. Comprende los municipios de Aldana, Contadero, Córdoba, Cuaspud Carlosama, Cumbal, Funes, Guachucal, Gualmatan, Iles, Ipiales, Potosí, Puerres y Pupiales, que abarcan un total de 4.894 kilómetros cuadrados.

Uno de los principales problemas que presentaron las instituciones educativas de los municipios de la subregión de Obando, es el bajo rendimiento que se viene presentando en los últimos años en las pruebas SABER 5° [4]. Además, no se cuenta con una información histórica de los resultados obtenidos por los estudiantes de esta región en las pruebas Saber 5°, integrados en un mercado de

datos que permita obtener resultados confiables y de calidad para la realización de análisis y soporte de decisiones para el mejoramiento de la calidad en estas instituciones.

En el departamento de Nariño, ni en la subregión de Obando, no se han desarrollado proyectos de investigación que almacenen y analicen los resultados de las pruebas Saber 5°, que presentan los estudiantes de educación básica de las instituciones educativas de la región.

OBJETIVOS

El objetivo general de este proyecto fue diseñar y construir un mercado de datos que almacene datos históricos, limpios y transformados de los resultados de los estudiantes de las Instituciones Educativas de los municipios de la subregión de Obando del departamento de Nariño, que presentaron las pruebas Saber 5° entre los años 2014 y 2016, que genere información de calidad para soportar la toma de decisiones acertadas de los directivos de las instituciones educativas y secretarías de educación de los municipios en pro del mejoramiento de la calidad de la educación básica y media de esta subregión.

Los objetivos específicos planteados en esta investigación para lograr el objetivo general fueron:

- Analizar y evaluar los informes y reportes generados, por diferentes dimensiones, con la herramienta Pentaho acerca de las pruebas Saber 5°.
- Aplicar técnicas de limpieza y transformación de datos al repositorio temporal de datos de las pruebas Saber 5°
- Apropiar el conocimiento sobre las competencias evaluadas en las pruebas Saber 5.
- Diseñar e implementar el modelo multidimensional en estrella del mercado de datos de las pruebas Saber 5°.
- Implementar el proceso de carga de datos al mercado de datos Saber 5° a partir del repositorio temporal de datos.
- Integrar el mercado de datos Saber 5° con la herramienta de inteligencia de negocios libre Pentaho.
- Integrar en una base de datos inicial la información de las pruebas Saber 5 ° de los estudiantes de las instituciones educativas de la subregión de Obando del departamento de Nariño, a partir de las bases de datos del ICFES de los años 2014 al 2016.

- Seleccionar de la base de datos inicial Saber 5°, los atributos socioeconómicos, académicos e institucionales de cada estudiante de las instituciones educativas de la subregión de Obando y almacenarlos en un repositorio temporal de datos.

JUSTIFICACIÓN

Este proyecto de investigación tuvo como objetivo construir un mercado de datos que almacene la información histórica de los estudiantes de las Instituciones Educativas de los municipios de la subregión de Obando, que presentaron las pruebas Saber 5° entre los años 2014 a 2016. Este mercado de datos, gestionado por una herramienta de Inteligencia de Negocios de software libre como lo es Pentaho, permitió realizar una vista dimensional de los datos con una máxima efectividad para ganar perspicacia y entender lo necesario para tomar optimas decisiones que mejoren la calidad de la educación en las instituciones de esta subregión.

Por otra parte, un mercado de datos de las pruebas Saber 5°, diseñado bajo un modelo multidimensional y que almacene información limpia y transformada, permitirá la distribución de los resultados del análisis en múltiples formatos y de forma inmediata.

En este sentido, este proyecto se convierte en el primero que se desarrolla en el departamento de Nariño y específicamente en la subregión de Obando, que busca generar información confiable y de calidad acerca de los resultados de las pruebas Saber 5° para soportar la toma de decisiones de los directivos de las instituciones educativas, de las secretarías de educación de los municipios de esta subregión e incluso del Ministerio de Educación Nacional (MEN), en la formulación de planes que permitan el mejoramiento de la calidad de la educación básica y media de la región.

Para las Instituciones Educativas y por ende las secretarías de educación de los municipios de la subregión de Obando, es importante tener información clasificada y de calidad para soportar la toma de decisiones en pro del mejoramiento de la calidad educativa, es así como se realizó el proyecto, el cual tuvo como objetivo el desarrollo de un mercado de datos para el análisis multidimensional, por medio de la metodología presentada por Kimball denominada Ciclo de Vida Dimensional del Negocio, una de las más utilizadas en este tipo de proyectos. Reconociendo la importancia de esta información, se pudo obtener resultados estadísticos a través de la herramienta libre Pentaho.

Este trabajo representa un aporte a la comunidad educativa y científica en el área de inteligencia de negocios, específicamente, en el descubrimiento de conocimiento en bases de datos; así como, a nivel formal y técnico, la carga y extracción de datos, consultas SQL y visualización de resultados, representarían una contribución en la investigación básica de la misma área. Fue importante que el desarrollo de esta investigación conto con el respaldo del Grupo de Investigación de Ingeniería de Sistemas de la Universidad de Nariño (GRIAS) y el producto obtenido sirva como respaldo para otras investigaciones.

ORGANIZACIÓN DE ESTE DOCUMENTO

Este documento está dividido en 7 capítulos principales nombrados de la siguiente manera: Introducción, Fundamentos teóricos, Conceptualización de pruebas saber 5°, Metodología, Implementación, Resultados y análisis, Conclusiones y recomendaciones.

En el capítulo 1, se presenta el planteamiento del problema, objetivo general y específicos, la justificación del trabajo, y las contribuciones de la investigación.

En el capítulo 2, se presenta los fundamentos teóricos en el cual se desarrolla esta investigación.

En el capítulo 3, se presenta la conceptualización de las pruebas saber 5°.

En el capítulo 4, se describe la metodología propuesta para el desarrollo de esta investigación.

En el capítulo 5, se presenta la implementación que se realizó en base a la metodología propuesta.

En el capítulo 6, se presenta los resultados que se obtuvo.

Por último, en el capítulo 7, se presenta las conclusiones y recomendaciones de la investigación.

1. FUNDAMENTOS TEÓRICOS

1.1. Data Warehouse

Este término acuñado por Bill Inmon, se traduce literalmente como Almacén de Datos. No obstante, si el Data Warehouse fuese exclusivamente un almacén de datos, los problemas seguirían siendo los mismos que en los Centros de Información.

La ventaja principal de este tipo de sistemas se basa en su concepto fundamental, la estructura de la información. Este concepto significa el almacenamiento de información homogénea y fiable, en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales. Según definió Bill Inmon, el Data Warehouse se caracteriza por ser:

Integrado: los datos almacenados en el Data Warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

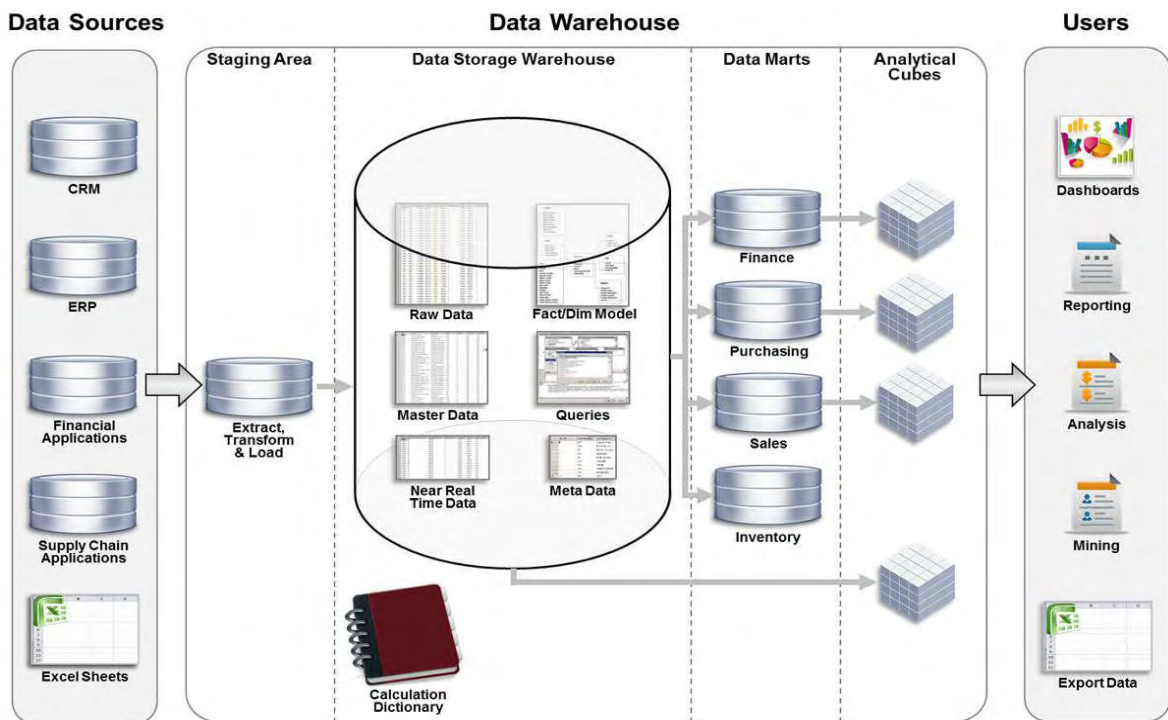
Temático: sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del Data Warehouse. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

Histórico: el tiempo es parte implícita de la información contenida en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

No volátil: el almacén de información de un Data Warehouse existe para ser leído, y no modificado. La información es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía [8].

Un Data Warehouse es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. La creación de un Data Warehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence. La ventaja principal de este tipo de bases de datos radica en las estructuras en las que se almacena la información (modelos de tablas en estrella, en copo de nieve, cubos relacionales... etc.) Este tipo de persistencia de la información es homogénea y fiable, y permite la consulta y el tratamiento jerarquizado de la misma (siempre en un entorno diferente a los sistemas operacionales) [9].

Figura 1. Estructura lógica de una Bodega de Datos.



Fuente: The Data Warehouse Group (<http://www.data-warehouse.com.au>).

1.2. Datamart

Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento.

Un Datamart puede ser alimentado desde los datos de un Data Warehouse, o integrar por sí mismo un compendio de distintas fuentes de información.

Por tanto, para crear el Datamart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP, como el propio Data Warehouse, o sobre una base de datos OLAP. La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento. De esta forma se pueden plantear dos tipos de datamarts:

Datamart OLAP: se basan en los populares cubos OLAP, que se construyen agregando, según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo, en función de la herramienta final que se utilice.

Datamart OLTP: pueden basarse en un simple extracto del Data Warehouse, no obstante, lo común es introducir mejoras en su rendimiento (las agregaciones y los filtrados suelen ser las operaciones más usuales) aprovechando las características particulares de cada área de la empresa. Las estructuras más comunes en este sentido son las tablas report, que vienen a ser fact-tables reducidas (que agregan las dimensiones oportunas), y las vistas materializadas, que se construyen con la misma estructura que las anteriores, pero con el objetivo de explotar la reescritura de queries (aunque sólo es posible en algunos SGBD avanzados, como Oracle).

Los Datamarts que están dotados con estas estructuras óptimas de análisis presentan las siguientes ventajas:

- Poco volumen de datos.
- Mayor rapidez de consulta.
- Consultas SQL y/o MDX sencillas.
- Validación directa de la información.

- Facilidad para la hostilización de los datos [10].

1.3. Diseño de un Datamart

Existen múltiples alternativas para el diseño de data marts entre los cuales se encuentran los aportes de Kimball y Rose [8], el trabajo de Malinowski y Zimányi [13] o la propuesta de Inmon [14]. La metodología de Ralph Kimball presenta un enfoque Bottom-Up (de menor a mayor o ascendente), lo cual se traduce en la implementación de data marts en áreas específicas de la organización, y poco a poco irlos integrando en un Data Warehouse Corporativo.

Rivadera en [12] describe los cuatro principios básicos que fundamentan el ciclo de vida de la metodología Kimball:

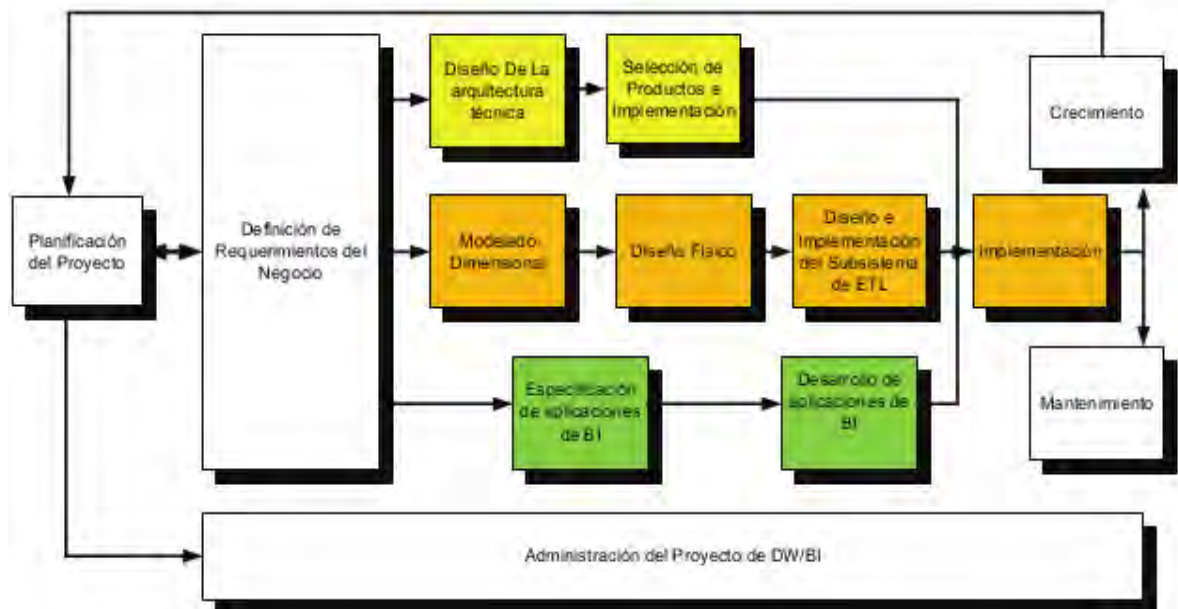
- Centrarse en el negocio: es fundamental identificar los requerimientos del negocio y su valor asociado.
- Construir una infraestructura de información adecuada: la herramienta diseñada debe ser fácil de usar, de alto rendimiento, integrada y que refleje los requerimientos del negocio.
- Realizar entregas en incrementos significativos: construir el DW en incrementos entregables; es una concepción similar a las metodologías ágiles de construcción de software.
- Ofrecer la solución completa: construir los elementos necesarios para entregar valor a los usuarios: un DW sólido (probado y accesible), herramientas de análisis y de consulta, generación de reportes, capacitación, soporte y documentación.

Esta metodología de desarrollo se compone de fases, como se muestra en la Figura 2, en la cual es importante destacar dos aspectos:

- 1) El rol central de la definición de requerimientos, que constituyen el soporte inicial de las actividades subsiguientes además de su influencia en la construcción del plan de proyecto.
- 2) Se identifican tres rutas o caminos enfocadas en diferentes áreas:
 - Tecnología (Camino Superior): involucra análisis de software y hardware.
 - Datos (Camino del medio): diseño e implementación del DW.

- Aplicaciones de Inteligencia de Negocios (Camino Inferior): diseño e implementación de aplicaciones para los usuarios finales.

Figura 2. Fases de la metodología Kimball.



Fuente: Ralph Kimball, The Data Warehouse Lifecycle Toolkit.

1.4. Análisis multidimensional

Hace referencia a las diferentes perspectivas que pueden ser usadas para analizar los datos almacenados en una bodega de datos; las tablas de dimensiones contienen los atributos que se quieren estudiar según el nivel de detalle. En un modelo dimensional se pueden encontrar los siguientes componentes [15] [16]:

- Dimensiones: representan las áreas temáticas, las líneas o los sujetos del negocio, brindan una línea base para estructurar la información mediante el uso de un grupo de atributos.
- Atributos: agrupación de elementos o ítems relacionados dentro de una dimensión que tienen el mismo nivel lógico
- Jerarquías: son representadas por un ordenamiento lógico dentro de la dimensión, se encuentran formadas por los diferentes tipos de relaciones entre los atributos de una misma dimensión
- Tablas de Hechos: un hecho es un indicador de negocio que da sentido al análisis de las dimensiones. Las tablas de hechos involucran indicadores

relacionados con un proceso específico del negocio y las llaves o claves de las dimensiones que intervienen en dicho proceso.

1.5. PostgreSQL

PostgreSQL es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD (Berkeley Software Distribución) y con su código fuente disponible libremente. Es el sistema de gestión de bases de datos de código abierto más potente del mercado y en sus últimas versiones no tiene nada que envidiarles a otras bases de datos comerciales.

PostgreSQL utiliza un modelo cliente/servidor y usa multiprocesos en vez de multi hilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando. [7]

1.6. Pentaho

La plataforma Open Source Pentaho Business Intelligence cubre muy amplias necesidades de Análisis de los Datos y de los Informes empresariales. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Eso hace que Pentaho es una solución muy flexible para cubrir una amplia gama de necesidades empresariales – tanto las típicas como las sofisticadas y específicas al negocio.

Los módulos de la plataforma Pentaho BI son:

Reporting: un módulo de los informes ofrece la solución adecuada a las necesidades de los usuarios. Pentaho Reporting es una solución basada en el proyecto JFreeReport y permite generar informes ágiles y de gran capacidad. Pentaho Reporting permite la distribución de los resultados del análisis en múltiples formatos todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. Los reportes Pentaho permiten también programación de tareas y ejecución automática de informes con una determinada periodicidad.

Análisis: Pentaho Análisis suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (pivot tables, crosstabs), generadas por Mondrian y JPivot, el usuario puede navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando

los campos de agregación. Los datos pueden ser representados en una forma de SVG o Flash, los dashboards widgets, o también integrados con los sistemas de minería de datos y los portales web (portlets). Además, con el Microsoft Excel Analysis Services, se puede analizar los datos dinámicos en Microsoft Excel (usando la conexión a OLAP server Mondrian).

Dashboards: Todos los componentes del módulo Pentaho Reporting y Pentaho Análisis pueden formar parte de un Dashboard. En Pentaho Dashboards es muy fácil incorporar una gran variedad en tipos de gráficos, tablas y velocímetros (dashboard widgets) e integrarlos con los Portlets JSP, en donde podrá visualizar informes, gráficos y análisis OLAP.

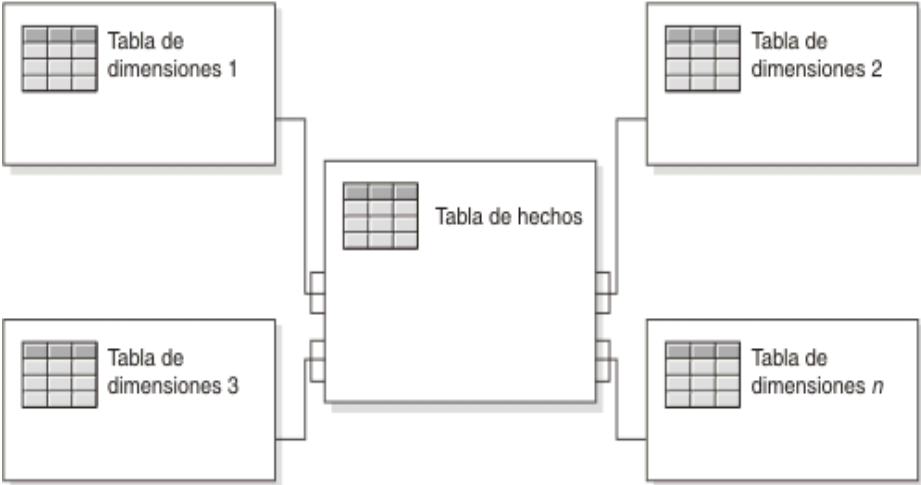
Data Mining: Análisis en Pentaho se realiza con una herramienta WeKa. Integración de Datos – se realiza con una herramienta Kettle ETL (Pentaho Data Integration) que permite implementar los procesos ETL. Últimamente Pentaho lanzó una nueva versión – PDI 3.0 – que marcó un gran paso adelante en OSBI ETL y que hizo Pentaho Data Integration una alternativa interesante para las herramientas comerciales. [11].

1.7. Modelo multidimensional en estrella

Un esquema de estrella es un tipo de esquema de base de datos relacional que consta de una sola tabla de hechos central rodeada de tablas de dimensiones.

En la siguiente figura se muestra un esquema de estrella con una sola tabla de hechos y cuatro tablas de dimensiones. Un esquema de estrella puede tener cualquier número de tablas de dimensiones. Las ramas situadas al final de los enlaces que conectan las tablas indican una relación de muchos a uno entre la tabla de hechos y cada tabla de dimensiones [18].

Figura 3. Esquema Tipo Estrella.



Fuente: Sinnexus.com. (2016). Modelo dimensional.

2. CONCEPTUALIZACIÓN DE PRUEBAS SABER 5°

2.1. Elementos conceptuales de las pruebas saber

El propósito principal de las pruebas SABER 3°, 5° y 9° es contribuir al mejoramiento de la calidad de la educación colombiana mediante la realización de evaluaciones aplicadas periódicamente para monitorear el desarrollo de las competencias básicas en los estudiantes de educación básica, como seguimiento de calidad del sistema educativo.

Las competencias en Lenguaje, Matemáticas, Ciencias Naturales y Competencias Ciudadanas son áreas, que a criterio de expertos se constituyen en los componentes principales para desarrollar en los estudiantes, las competencias básicas para la vida y que les permiten alcanzar un desempeño eficaz y eficiente dentro de la sociedad [1].

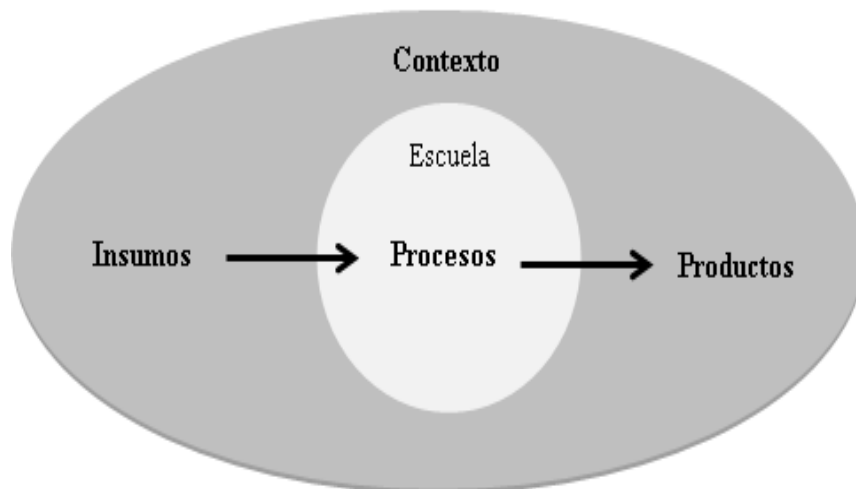
Los resultados de estas evaluaciones y el análisis de los factores asociados que inciden en los desempeños de los estudiantes, permiten que los establecimientos educativos, las secretarías de educación, el Ministerio de Educación Nacional y la sociedad en general identifiquen las destrezas, habilidades y valores que los estudiantes colombianos desarrollan durante la trayectoria escolar, independientemente de su procedencia, condiciones sociales, económicas y culturales, con lo cual, se puedan definir planes de mejoramiento en sus respectivos ámbitos de actuación [3].

Su carácter periódico posibilita, además, valorar cuáles han sido los avances en un determinado lapso y establecer el impacto de programas y acciones específicas de mejoramiento [12].

Sumado a esto, las pruebas permiten realizar estudios sobre las variables que tienen más influencia en el rendimiento académico y el ICFES con el fin de mejorar la calidad de la educación del país extendió el proceso de evaluación e inició el estudio de los factores asociados al rendimiento escolar, para este fin se aplicó un cuestionario de prueba en el año 2012 para recoger información sobre el contexto de los estudiantes, sus familias y la institución educativa. Para este estudio el ICFES implementó el modelo Contexto, Insumos, Procesos y Productos (CIPP) el cual permite seleccionar variables de un contexto particular para

describir de forma clara y sencilla como éstas influyen en el rendimiento académico. Ver figura 4.

Figura 4. Modelo CIPP.



Fuente: Marco de factores asociados Saber 3°, 5° y 9° de 2016.

En el contexto se recogen todas las variables externas a la institución educativa que y se caracterizan por los aspectos: social, económico, político, cultural y familiar. Dentro de estos se incluyen la ubicación geográfica de la institución educativa, así como la modalidad, dependencia administrativa y tamaño del establecimiento. Otros aspectos relacionados con las características personales de los estudiantes y de sus hogares como el nivel socioeconómico, el género, la motivación y el auto concepto académico también hacen parte de esta categoría.

Los Insumos hacen referencia a los recursos con los que se cuenta en la institución educativa e incluye el historial académico de los estudiantes como un factor determinante del desempeño de los mismos. Dentro de los indicadores de infraestructura escolar están los servicios básicos, el acceso a computadoras y la conexión a internet, entre otras, el tiempo efectivo de aprendizaje y las estrategias de clasificación de estudiantes al interior de las sedes; las variables asociadas a los antecedentes escolares como la asistencia a educación preescolar y la repetición de grado también se incluyen en esta categoría.

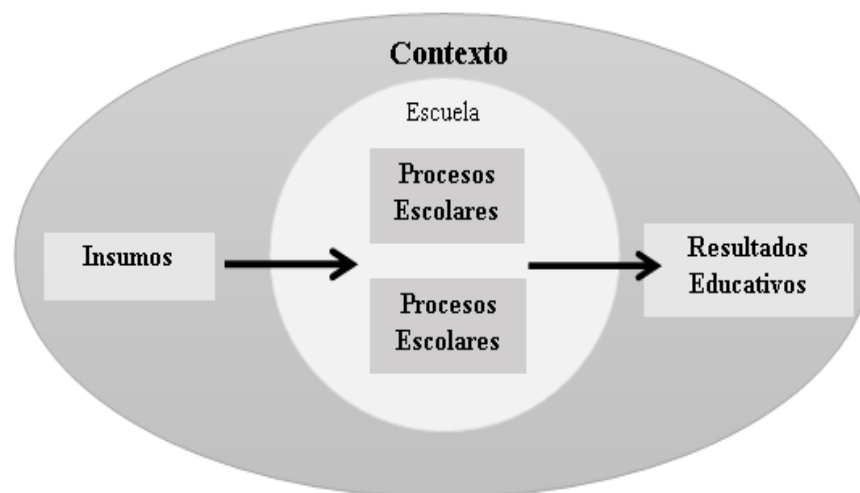
En los Procesos se tiene en cuenta las actividades y estrategias que se implementan en las instituciones educativas para que los estudiantes apropien los conocimientos y logren desarrollar las competencias que se establecen desde el MEN. Además, se encuentran aquellas variables que permiten caracterizar el clima al interior del aula y de la institución educativa, la gestión de los directores y la satisfacción de los docentes y de todo el personal de la institución.

Finalmente, los Productos son los resultados de las actividades empleadas en las instituciones y las políticas públicas, uno de esos resultados es el desarrollo cognitivo, social, emocional y ciudadano que conllevan a formar una persona integral.

El modelo CIPP cumple cuatro criterios generales que permiten hacer un juicio valorativo; es completo puesto que incluye todas las categorías de los factores asociados al aprendizaje y la influencia de estas en el proceso educativo. Es claro, porque presenta resultados que se pueden entender a partir cualquier ámbito desde especialistas en investigación educativa, así como tomadores de decisiones, docentes y rectores. Presenta direccionalidad de las relaciones que se espera encontrar a partir del planteamiento de hipótesis sobre los factores asociados al aprendizaje. Por último, el modelo es flexible y permite incluir variables que son relevantes para el aprendizaje y desarrollo integral de los estudiantes, esto debido a que se pueden presentar cambios o transformaciones socioculturales [4].

2.1.1. Factores asociados. Como se mencionó anteriormente el marco de factores asociados sigue el modelo CIPP, como se muestra en la figura 5, y dentro de cada una de las categorías incluidas en él se definen unas variables que han sido conceptualizadas y analizadas a nivel mundial y pueden usarse de forma dinámica para responder a políticas públicas o adaptarse a cualquier contexto a nivel nacional, regional y local [4]. Se definen a continuación el Marco de factores asociados Saber 3°, 5° y 9°:

Figura 5. Marco de Factores Asociados.



Fuente: Marco de factores asociados Saber 3°, 5° y 9° de 2016.

2.1.1.1. Contexto. Los factores asociados concentran las variables más cercanas al proceso educativo porque son estas las que caracterizan a los estudiantes e instituciones educativas, algunas de ellas son:

Características de los Estudiantes

Los aspectos sociodemográficos, económicos y culturales de los estudiantes y sus familias son variables importantes a la hora de explicar los resultados escolares. Estos factores a pesar de influir en el aprendizaje no se pueden modificar puesto que no dependen las instituciones educativas. Las variables que se consideran de mayor seguimiento dentro del marco de factores asociados son [20] [19]:

- **Género:** Se evidencia una brecha de desigualdad en el rendimiento académico de los niños y niñas puesto que se ven reflejados los estereotipos de género que se marcan desde la familia y la sociedad, por ejemplo, la gran desventaja que se muestra en las niñas frente al aprendizaje de las matemáticas y a los niños en lectura [19] [23]; [21] [22].
- **Edad:** Esta variable se relaciona con el grado que cursa el estudiante; permite identificar aspectos como la repetición de año, la extra edad, la desescolarización parcial o deserción escolar.

Según la ley general de educación un estudiante se encuentra en extra edad si supera en dos años la edad promedio del grado cursado; por tanto, la escolarización es obligatoria entre los 5 y 15 años de edad, desde primero a noveno grado [4].

- **Trabajo Infantil:** Es una variable que afecta de manera negativa al rendimiento escolar ya que las condiciones físicas de los niños y niñas disminuyen, pues no están preparados para asumir tareas laborales pesadas.

Se debe diferenciar entre el trabajo infantil y la colaboración en tareas hogareñas puesto que el primero afecta negativamente el rendimiento escolar y el segundo muestra una relación positiva [24] [19].

- **Nivel Socioeconómico Familiar:** La variable socioeconómica es explicativa frente al desempeño de los estudiantes en las Pruebas Saber, puesto que es una limitante para su desarrollo. Las condiciones del entorno donde viven, el nivel educativo de los padres, la situación laboral, la disposición de artefactos

tecnológicos, entre otras, son algunos aspectos de interés ya que son un aliciente importante para el rendimiento escolar [25] [63].

- **Violencia en el entorno del hogar:** Un entorno violento tiene influencias negativas en cuanto al rendimiento escolar y la formación ciudadana. Para poder determinar el efecto que tiene la violencia en el entorno familiar y en el desempeño académico se debe hacer un estudio de la frecuencia con que ocurren dichas situaciones y en qué tipo de violencia se está incurriendo [26].
- **Involucramiento Parental:** El acompañamiento y participación de los padres en el proceso de formación académica beneficia el desarrollo cognitivo y no cognitivo de los hijos. Crea actitudes de motivación frente al aprendizaje, reflejadas en los resultados escolares, se fortalece el desarrollo motivacional y ayuda a crear hábitos de estudio en familia [27].
- **Pertenencia a un Pueblo Originario o Etnia:** El desafío que hay en las instituciones educativas es el de incluir y atender a estudiantes que pertenecen a alguna etnia en particular; ya que estos presentan condiciones socioeconómicas bajas y de difícil acceso a la educación. Por esto, los establecimientos educativos deben apropiarse sus currículos para la flexibilización del aprendizaje y las necesidades que requeridas por los estudiantes [64].
- **Necesidades Educativas Especiales:** La política de una “Educación para todos” planteada por la UNESCO deja ver la importancia de tener en cuenta esta variable NEE dentro de los currículos de los establecimientos educativos del país, ya que incide significativamente en el desempeño académico de los estudiantes. Se deben distinguir entre necesidades educativas cognitivas y físicas, dado que en las primeras se debe buscar fortalecer y alcanzar los logros mínimos planteados para cada grado, mientras que las necesidades físicas se asocian más a la adecuación de espacios físicos que le permita la accesibilidad al establecimiento [28].
- **Auto-Concepto Académico:** Es entendido como las creencias que tienen los estudiantes sobre sus capacidades para hacer bien sus labores académicas y obtener buenos resultados. La confianza que tenga cada estudiante sobre su desempeño afecta de manera positiva o negativa sus resultados; esto se evidencia en el cumplimiento de tareas, la participación en clases y la asistencia a las mismas.

En la formación del auto-concepto académico se tiene en cuenta aspectos como la comparación personal, la comparación social, la percepción de los padres, profesores y compañeros [29].

- **Motivación:** Esta variable muestra una estrecha relación con el desempeño académico de los estudiantes es así como se proponen dos tipos de motivación: la intrínseca cuando se refiere a la motivación causada por gusto e interés propio y la extrínseca cuando es causada por algún incentivo en particular [30]. En el marco conceptual de los cuestionarios PISA, la motivación es reconocida como parte de los comportamientos o actitudes positivas frente al aprendizaje [31] [32] [33] y cuando se logra controlar la variable socioeconómica la motivación intrínseca tiene resultados positivos frente al aprendizaje.

Además, la motivación en el ámbito educativo se debe entender como la manera en que las relaciones entre los estudiantes y sus profesores entran en juego para facilitarla o limitarla con el fin de lograr un aprendizaje significativo, razón por la cual el profesor debe favorecer la denominada motivación intrínseca, que lleva a la curiosidad y el descubrimiento de lo nuevo, a diferencia de la motivación extrínseca en la que el estudiante es movido por otros para realizar diferentes actividades [34].

- **Estrategias de Aprendizaje:** Las estrategias de aprendizaje son las que facilitan la adquisición del conocimiento y la adquisición de nuevas habilidades.

Existen diferentes estrategias de aprendizaje, pero se intensifican los estudios en las tradicionales y las cooperativas. Caracterizándose las primeras por el uso de la memorización como recurso de aprendizaje y las segundas por la creación colectiva del conocimiento [35].

Características de las Escuelas

En esta sección se analizan aquellas variables que hacen referencia a las instituciones educativas. Dichas características están referenciadas en los cuestionarios que se aplican a los estudiantes al momento de presentar las Pruebas Saber. Entre las variables más relevantes están:

- **Zona en la que se Ubica la Institución Educativa (Urbana o Rural):** Sobre esta variable se han realizado estudios que evidencian que las instituciones que

se encuentran más aisladas poseen menos posibilidades de contar con docentes idóneos, materiales y recursos tecnológicos, apoyo técnico y se da una mayor deserción estudiantil, principalmente por la falta de recursos económicos de los padres que en muchas ocasiones emplean a sus hijos para realizar las labores del campo. Además de esto, en diversos establecimientos que se encuentran en las zonas rurales existen modos de organización diferentes pues es común encontrar escuelas; multigrado o unidocente [4].

- **Dependencia Administrativa de la Escuela:** La influencia ejercida por el tipo de dependencia es una de las variables que cobra mayor importancia dentro de las características de las escuelas, ya que algunos autores consideran que los entes privados exigen por parte de los directivos y docentes el mejoramiento continuo en los aprendizajes de los estudiantes por la presión ejercida por los padres [36]. Sin embargo, en la práctica se ha encontrado que dicha afirmación en pocos casos se cumple, indicando que la educación privada por sí sola no asegura mejores aprendizajes y oportunidades para los estudiantes [37].
- **Tamaño de la Institución Educativa y el Aula de Clases:** Frente a este factor existen contraposiciones de diversos autores frente al tamaño de la institución y el número de estudiantes en el aula; se afirma que la cantidad de estudiantes atendidos en la institución educativa influye en el tipo de enseñanza y en el proceso de aprendizaje de los estudiantes [38] [39], sin embargo también se enfoca más en el tamaño relativo de la institución afirmando que el número de docentes por estudiante explica de mejor manera los efectos en el apoyo del proceso de enseñanza [40] [41].
- **Nivel Socioeconómico Promedio por Aula y Establecimiento:** Este factor se puede determinar a partir del estrato de la institución y toma importancia al identificar el efecto de los padres en cada estudiante, el cual es beneficioso en el desarrollo escolar pues se da la heterogeneidad social y académica de los estudiantes, a nivel de aula y escuela, además impacta el desarrollo del currículum, la calidad de las interacciones y la capacidad del profesor para desarrollar procesos de enseñanza efectivos [42].
- **Violencia en el Entorno de la Institución Educativa:** Esta característica evidencia las situaciones de conflicto y por ende de procesos de marginación al interior de las instituciones educativas. Las repercusiones al respecto recaen en principalmente en los estudiantes, frustrando su adecuado desarrollo académico y posteriormente los bajos desempeños en Pruebas Saber. [43] [64].

2.1.1.2. Insumos. Los insumos o recursos educativos con los que cuenta la institución educativa son una parte importante en los resultados; teniendo en cuenta que con estos los estudiantes obtienen un aprendizaje significativo que fortalece su nivel académico. A continuación, se presentan aspectos que se relacionan con esta categoría:

- **Antecedentes Escolares:** Es importante reconocer que los estudiantes traen aprendizajes y conocimientos previos que influyen de manera directa en el rendimiento académico. Factores asociados como la participación en preescolar y la repetición de grado son antecedentes que traen consigo los estudiantes y se consideran favorables para la adquisición de nuevas competencias en la formación académica [4].
- **Asistencia a Educación Preescolar:** La educación preescolar tiene efectos positivos sobre los resultados escolares y en otros ámbitos como la reducción de la pobreza y la movilidad social; por tal razón la asistencia al preescolar es favorable para el desarrollo de la sociedad en general [44].
- **Repetición de Grado:** Este tema ha sido estudiado por diferentes autores quienes indican que la repetición de año tiene efectos negativos a nivel académico y emocional a través del tiempo. Cuanto menor sea el grado de repetición mayores serán los efectos o consecuencias negativas para el aprendizaje [46] [47]. En Latinoamérica, la repitencia de grado en la institución educativa ha sido identificada como la variable que más influencia en los logros de aprendizaje, luego del nivel socioeconómico de los estudiantes, lo cual refleja un impacto negativo como medida efectiva para mejorar el desempeño académico [64]. En la actualidad Colombia es uno de los países de Latinoamérica con las tasas de repitencia más bajas en educación primaria; sin embargo, al considerar las consecuencias de esto, resulta conveniente revisar con mayor detalle las políticas del sistema educativo colombiano [24].
- **Infraestructura y Equipamiento Escolar:** La relación que existe entre la infraestructura y el rendimiento académico presenta una relación positiva ya que; disponer de una biblioteca, de computadores o tecnologías de información y comunicación, de espacios recreativos, de salones de clase adecuados para todos los estudiantes y de servicios sanitarios adecuados, le da la posibilidad a los estudiantes de contar con ambientes propicios que generan mayor interés y motivación por el aprendizaje [64].

- **Calificación Docente y Conocimiento Profesional:** Un buen docente se caracteriza por tres aspectos que lo hacen idóneo en su desempeño laboral: los criterios de selección inicial, la calidad de instrucción y la capacitación continua.
- **Materiales Educativos:** Hace referencia a los recursos o herramientas que permiten la profundización de los conceptos y la apropiación del conocimiento tanto de estudiantes como docentes.

Algunos elementos que hacen parte de los materiales educativos son los artefactos tecnológicos, los libros de texto, la disponibilidad de internet, los útiles escolares, entre otros. El uso y aprovechamiento que se les dé a estos recursos en el aula de clases durante el desarrollo de las sesiones pedagógicas define en gran medida los resultados esperados frente al logro educativo [4].

- **Programa de Útiles Escolares Gratuitos:** Si bien es cierto este tipo de programas son de vital importancia para apoyar los procesos de aprendizaje de los estudiantes, ya que les permite tener acceso y disponibilidad de herramientas educativas para la adquisición de conocimientos, también se sabe que estos programas tienen criterios de focalización (colegios o estudiantes más desventajados socioeconómicamente) sin embargo los materiales que entregan a los establecimientos educativos muchas veces no son necesariamente los que se requieren para orientar el aprendizaje, pues como se mencionó anteriormente no se cuenta con la instrucción adecuada para que el uso adecuado de dichas herramientas [49].
- **Tiempo Efectivo Dedicado al Aprendizaje:** Este componente hace referencia al tiempo que se dedica exclusivamente al aprendizaje haciendo diferencia entre las horas de estudio de los estudiantes y el tipo de jornada al que asisten. Otros factores que permiten identificar el tiempo efectivo dedicado al aprendizaje como son la inasistencia escolar y la puntualidad por parte de estudiantes y docentes, la frecuencia con la que ocurren interrupciones durante la clase, debido a un ambiente de indisciplina nocivo para el aprendizaje o a factores de distracción externos al control del docente como ruido o condiciones climáticas inapropiadas, y el manejo y organización del tiempo por parte del docente [50].

2.1.1.3. Procesos. Los procesos que ocurren en la institución educativa y en los salones de clase representan el núcleo de la labor educativa, puesto que a través de las interacciones cotidianas entre docentes y estudiantes se promueve el aprendizaje [4].

Los constructos que hacen parte de los procesos educativos se definen a continuación:

- **Liderazgo Educativo:** Esta es una de las variables que más influye en el aprendizaje de los estudiantes dado el papel que juega el desarrollo de programas, proyectos y prácticas educativas. Los roles de liderazgo educativo orientan principalmente al logro del aprendizaje y se focalizan en los aspectos de la gestión instruccional. El liderazgo instruccional es un factor importante para orientar las metas y focalizar los agentes internos hacia los resultados académicos en los procesos de enseñanza. Algunos de los procesos escolares transformadores que deben ser tomados en cuenta por las personas a cargo del liderazgo educativo son: empoderamiento de los profesores, fortalecimiento de la visión institucional y de objetivos compartidos, y capacidades de rendición de cuentas [51].
- **Desarrollo y Colaboración Profesional Para Mejorar la Enseñanza:** El trabajo cooperativo y colaborativo entre maestros permite mejorar el aprendizaje y el desempeño escolar; el intercambio de experiencias sobre las prácticas de aula, conlleva a mejorar las propias y contribuye a desarrollar habilidades en los estudiantes.

Además de esto, el trabajo colaborativo permite fortalecer la planificación, el desarrollo, la evaluación y la retroalimentación para mejorar las prácticas educativas [52] [19] [33].

- **Clima Escolar:** Hace alusión al conjunto de normas y acuerdos que se establecen en la institución educativa para propiciar un ambiente acorde a las necesidades de los estudiantes y garantizar un aprendizaje efectivo. Un clima laboral y escolar positivo se asocia con los logros académicos [33].

Algunos constructos que se relacionan con esta variable son:

- a) **Presencia Efectiva de Normas y Acuerdos:** Un ambiente de aprendizaje armonioso, basado en el respeto con normas claras y el cumplimiento de las mismas, potencia el aprendizaje estudiantil; contrario a esto, cuando se presentan ambientes hostiles, de autoritarismo e irrespeto por parte del docente o los estudiantes, se genera indisciplina y pérdida de tiempo valioso que podría dedicarse a la adquisición de conocimientos [33].

b) Relaciones Interpersonales: Entre los aspectos a tener en cuenta en las relaciones interpersonales están la comunicación efectiva y asertiva entre los actores educativos, así como las relaciones basadas en la confianza. Es impotente hacer énfasis en la actitud y disposición que tienen los docentes para resolver las dudas de sus estudiantes, la actitud de los estudiantes frente a los procesos de enseñanza y el tipo de comunicación que se de de en clase [53].

- **Prácticas de Enseñanza Docente:** Este aspecto hace énfasis en la calidad instruccional del profesor en donde se evidencia la interacción entre este y sus estudiantes, generando una relación positiva si se cuenta con un conjunto de habilidades por parte del docente, entre las cuales están en primer lugar el desarrollo de conceptos, la estimulación y el desarrollo de habilidades de pensamiento superior, la integración de conceptos y de conocimientos previos, la producción propia y las relaciones y aplicaciones al mundo real. En segundo lugar, la calidad de la retroalimentación dada por parte del docente a los estudiantes, donde el profesor debe desarrollar bases claras para que el estudiante logre avanzar en su aprendizaje, que haga seguimiento constante, que promueva la metacognición y que reafirme positivamente a los estudiantes. Y, en tercer lugar, la modelación lingüística, que se refiere a la capacidad del profesor para extender y desarrollar constantemente el lenguaje en los estudiantes, a través de conversaciones frecuentes, preguntas abiertas y el uso de lenguaje avanzado, entre otros [60].

2.1.1.4. Resultados Educativos. Actualmente la expectativa de los sistemas educativos, frente al objetivo de promover el desarrollo integral de los estudiantes, se ha convertido en la base fundamental para propender por un buen desempeño académico; partiendo del aprendizaje en las disciplinas académicas, el desarrollo de competencias ciudadanas y el bienestar subjetivo de los estudiantes. Por esta razón es importante describir cada una de ellas como se hace a continuación [60].

- **Aprendizaje en Disciplinas Académicas:** Según lo establecido en los estándares básicos de competencias es de vital importancia tener un criterio claro y público que permita juzgar si un estudiante, una institución o el sistema educativo en su conjunto cumplen con unas expectativas comunes de calidad, a esto se define estándar y partiendo de esta definición cobra importancia la medición de los aprendizajes en las distintas disciplinas académicas de las cuales no solo hacen parte las áreas básicas como matemáticas y lenguaje,

sino también aquellas que forman parte del desarrollo integral del estudiante y su formación como ciudadanos [55].

- **Ciudadanía:** La formación ciudadana hace referencia, de forma general, a aprender a vivir con otros y a participar activamente de una sociedad democrática. En este sentido, este factor asociado se debe tener en cuenta como primordial en la formación académica y ciudadana de los estudiantes, siendo las instituciones educativas el lugar propicio para la socialización y el desarrollo de habilidades y actitudes en relación a la vida en democracia, el respeto por los otros, el cumplimiento de las normas acordadas, participar y expresar su opinión, entre otros aspectos [57] [56].

2.1.2. Estructura y alineación de las pruebas saber. A partir del segundo semestre del año 2014 las pruebas fueron alineadas para establecer comparativos con los diferentes exámenes que se aplican como son las Pruebas Saber 3°, 5°, 9°, Saber 11 y Saber Pro. Esto con el fin de monitorear el progreso que han tenido los estudiantes después de aplicar la prueba en un determinado grado, por ejemplo, un estudiante de primaria que presentó la prueba Saber 3° después de dos años aplica la prueba Saber 5° con los resultados obtenidos de las dos pruebas se puede establecer un comparativo y hacer un análisis del proceso de formación que alcanzó en la Básica Primaria [58].

Actualmente, el Ministerio de Educación Nacional (MEN) concibe el objetivo de la educación como el desarrollo de determinadas competencias y, en consecuencia, a estas como el objeto de la evaluación. Dentro de las diferentes competencias que pueden desarrollarse a lo largo del proceso educativo hay una categoría que merece especial atención: la de las competencias genéricas, entendidas como aquellas que resultan indispensables para el desempeño social, laboral y cívico de todo ciudadano, independientemente de su oficio o profesión. Contrastan con las competencias (no genéricas) propias de oficios o actividades laborales particulares, que resultan de un entrenamiento especializado [58].

Una competencia es más que conocimientos y destrezas. Involucra la habilidad de enfrentar demandas complejas, apoyándose y movilizando recursos psicosociales (incluyendo destrezas y actitudes) en un contexto en particular. Por ejemplo, la habilidad para comunicarse de manera asertiva a través del buen uso del lenguaje, de actitudes positivas con otras personas y buenas prácticas comunicativas en su entorno, son aspectos que evidencian la movilización de competencias [59].

En este documento, se entienden las competencias como un conjunto identificable y evaluable de conocimientos, actitudes, valores y habilidades relacionadas entre sí, que permiten desempeños satisfactorios en situaciones reales y en contextos específicos [61]. En este sentido, una persona demuestra que es competente en la acción y no con la repetición de un saber determinado; en otras palabras, se evidencia la competencia cuando el conjunto de saberes se proyecta en acciones concretas que demandan su ejecución consciente; por lo tanto, las competencias se manifiestan en los desempeños que tiene el estudiante en situaciones específicas. Las competencias reconocen diversos grados de desempeño y de logros, expresados mediante indicadores, que permiten identificar los diferentes momentos o niveles de logro que constituyen una competencia determinada. Las competencias son dinámicas y se transforman de acuerdo con las experiencias del individuo y con los procesos de aprendizaje [62].

El ICFES para el año 2009 diseñó las Pruebas Saber 3°, 5° y 9° de tal forma que garanticen evaluaciones censales para un periodo de doce años con el fin de visualizar en los resultados la evolución que ha tenido la educación en Colombia. Estas pruebas evalúan las competencias en Matemáticas, Lenguaje y Ciencias Naturales; cabe resaltar que éstas no permiten evaluar la totalidad de las competencias que deben adquirir los estudiantes durante su estancia en la escuela, pero si son un aliciente para que ellos continúen su formación profesional, laboral y social a lo largo de toda su vida [65].

En la tabla 2 se evidencia la conformación de las Pruebas Saber 2009 teniendo en cuenta las competencias en Lenguaje, Matemáticas y Ciencias Naturales.

Tabla 2. Procesos de las competencias genéricas. Lenguaje Matemáticas Ciencias.

Lenguaje	Matemáticas	Ciencias Naturales
Lectura Escritura	Razonamiento y argumentación Comunicación, representación y modelación Planteamiento y resolución de problemas	Uso comprensivo del conocimiento científico Explicación de fenómenos Indagación

Fuente: Guía de lineamientos generales de Pruebas Saber 2009.

Además de esto, en cada una de ellas se evalúan los componentes que son los ejes verticales con los Estándares Básicos de Competencias y que permiten evidenciar las fortalezas y debilidades que tienen los estudiantes como se muestra en la tabla 3.

Tabla 3. Componentes de las competencias genéricas.

Lenguaje	Matemáticas	Ciencias Naturales
Semántica	Numérico – Variacional	Entorno vivo
Sintaxis	Geométrico – Métrico	Entorno físico
Pragmática	Aleatorio	Ciencia, tecnología y sociedad (CTS)

Fuente. Guía de lineamientos generales de Pruebas Saber 2009.

La Prueba comprende de preguntas de selección múltiple con única respuesta, estas presentan un enunciado con cuatro opciones de respuesta A, B, C y D y sólo una de ellas es la correcta, el número de preguntas que contesta cada estudiante y en particular los de grado quinto son: 36 para Lenguaje, 48 en Matemáticas y 48 para ciencias Naturales.

En el Plan Nacional Decenal de Educación 2006-2016, el ICFES ha avanzado en la alineación del Sistema Nacional de Evaluación Externa Estandarizada (SNEE), a través de la reestructuración de los exámenes: en 2009 con un nuevo diseño de Saber 3°, 5° y 9°; en 2010 con el rediseño de Saber Pro; en 2014 con los cambios en Saber 11°. La alineación posibilita la comparación de los resultados en distintos niveles educativos, ya que los diferentes exámenes evalúan unas mismas competencias en algunas de las áreas que los conforman, a saber, las competencias genéricas [2].

En el caso particular del examen Saber 5°, que se aplica desde el segundo semestre del 2014, la alineación consistió en la introducción de una prueba de competencias ciudadanas la cual presenta situaciones de análisis que se relacionan con su entorno más cercano, es decir, el aula, el colegio, la familia y el barrio, con menor complejidad y en un lenguaje más sencillo que las de grados 9° [2].

3. METODOLOGÍA

Esta investigación involucra la construcción de un mercado de datos y utiliza la metodología propuesta por Kimball denominada Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle) [17], una de las más utilizadas en proyectos de inteligencia de negocios (BI).

Esta metodología contempla seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación, que se describen a continuación.

3.1. Planificación

Determinar el propósito del proyecto, sus objetivos y el alcance. Incluye las acciones típicas de un plan de proyecto:

- Registrar los requerimientos del negocio.
- Identificar las tareas.
- Programar las tareas.
- Planificar el uso de los recursos.
- Asignar la carga de trabajo a los recursos.

3.2. Análisis de requerimientos

Se debe conocer tanto como se pueda del negocio, implica revisar informes, rastrear documentos, dominar los términos y la terminología del negocio.

3.3. Modelado dimensional

Proceso dinámico e iterativo. Inicia con un modelo dimensional de alto nivel generado a partir de los procesos priorizados de la matriz de requerimientos.

- Establecer el nivel de granularidad.
- Elegir las dimensiones.
- Identificar las tablas de hechos y medidas.
- Modelo gráfico de alto nivel.

- Identificación de atributos de dimensiones y tablas de hechos.
- Implementar el modelo dimensional detallado.
- Prueba del modelo.
- Revisión y validación del modelo.

3.4. Diseño físico

Se eligen las estructuras que soporten el diseño lógico; busca dar respuesta a preguntas como las siguientes:

- ¿Cómo puede determinar cuán grande será el sistema?
- ¿Cuáles son los factores de uso que llevarán a una configuración más grande y más compleja?
- ¿Cómo se debe configurar el sistema?
- ¿Cuánta memoria y servidores se necesitan?
- ¿Qué tipo de almacenamiento y procesadores son requeridos?
- ¿Cómo instalar el software en los servidores de desarrollo, prueba y producción?

3.5. Diseño del sistema de extracción, transformación y carga (ETL)

Busca alimentar el Data Warehouse; implica extraer los datos de los sistemas de origen, mejorar su calidad y consistencia y finalmente cargar los datos en el Datamart.

3.6. Especificación y desarrollo de aplicaciones de inteligencia de negocios bi (del inglés business intelligence)

Permitir a los usuarios finales acceso al Data warehouse de forma simple y estructurada a través de aplicaciones de inteligencia de negocios como informes y aplicaciones de análisis.

4. IMPLEMENTACIÓN

Para el desarrollo del proyecto se siguieron las fases propuestas por Kimball, en lo que denominó Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle). Las tareas realizadas en cada una de las fases son descritas a continuación.

4.1. Análisis de requerimientos

El interés de este proyecto se centró en construir un Datamart de las pruebas Saber 5° y de esta forma contar con una herramienta de inteligencia de negocios que permita desarrollar análisis multidimensional y que genera otros beneficios a los directivos de las Instituciones Educativas de la Subregión de Obando y de las Secretarías de Educación de estos Municipios. Por esta razón se necesitaba tener:

- Disponer de un repositorio de los resultados de las pruebas Saber 5° para el periodo comprendido entre 2014 a 2016.
- Analizar el desempeño histórico de los estudiantes en las pruebas Saber 5| para el periodo comprendido entre el 2014 a 2016

4.2. Modelado dimensional

Para la construcción del Datamart se contó inicialmente con datos de 2'183.435 estudiantes elegidos de manera aleatoria de las diferentes Instituciones académicas y un total de 65 atributos, que en términos generales pueden ser agrupados en las siguientes categorías:

- Características generales del estudiante.
- Características de los municipios.
- Características de la Instituciones Educativas.
- Características de la prueba.

Tabla 4. Dimensiones y granularidad del Datamart.

Dimensión	Granularidad
Estudiantes	Estudiante
Instituciones	Establecimientos de educación
Municipio	Municipios (subregión de Obando)

Prueba	Prueba (evaluada por áreas del conocimiento)
--------	----------------------------------------------

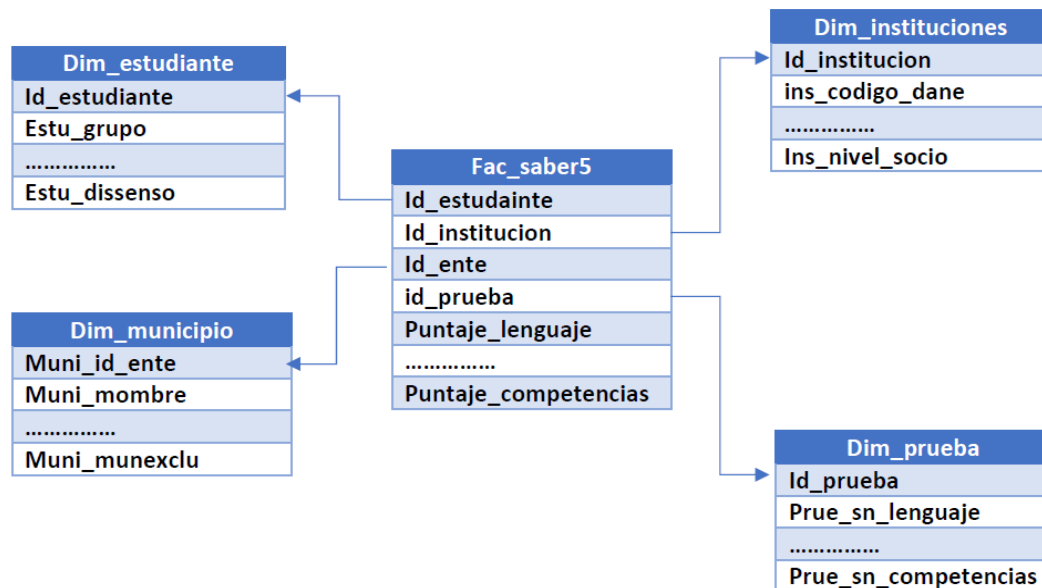
Fuente: Esta investigación.

Se tiene una tabla de hechos (fact_saber5) que tiene como granulo el estudiante, y se identificaron las siguientes medidas sobre los hechos:

- ¿Cuántos estudiantes presentaron las pruebas Saber 5° entre el 2014 al 2016?
- ¿Cuántos estudiantes presentaron cada una de las pruebas de las ciencias básicas del saber (Lenguaje, Matemáticas, Ciencias y competencias ciudadanas) entre el 2014 al 2016?
- ¿Cuál es el promedio del puntaje obtenido en cada una de las pruebas básicas del saber (Lenguaje, Matemáticas, Ciencias y competencias ciudadanas) entre el 2014 al 2016?
- ¿Cuántos estudiantes obtuvieron un desempeño por debajo de la media en la prueba?

Para el diseño se realizó el modelo multidimensional asociado al Datamart, se decidió usar un esquema tipo estrella, el cual se presenta en la Figura 6.

Figura 6. Modelo Multidimensional tipo estrella.



Fuente: Esta investigación.

En las Tablas 5 a 9 se detalla el diccionario de datos asociado a cada una de las tablas del modelo multidimensional.

4.3. Diccionario de datos del Datamart

Tabla 5. Diccionario de Datos de la Dimensión Estudiantes.

Nombre del Atributo	Tipo de dato	Descripción
id_estudiante	SERIAL	llave primaria
estu_consecutivo	TEXT	identificador del estudiante
estu_grupo	TEXT	grupo o salón al que pertenece el estudiante
estu_n	TEXT	matricula del grado al que pertenece el estudiante
estu_estrato	TEXT	estrato al que pertenece el establecimiento
estu_grado	TEXT	grado del estudiante
estusexo	TEXT	sexo del estudiante
estu_disenso	TEXT	marca de discapacidad cognitiva

Fuente: Esta investigación.

Tabla 6. Diccionario de Datos de la Dimensión Instituciones.

Nombre del Atributo	Tipo de dato	Descripción
id_institucion	SERIAL	llave primaria
ins_codigo_dane	TEXT	Código DANE del establecimiento educativo
ins_municipio	TEXT	Código DANE del municipio al que pertenece el establecimiento educativo
ins_id_ente	TEXT	id de la entidad territorial
ins_nombre	TEXT	Nombre del establecimiento educativo reportado en el DUE
ins_zona	TEXT	Zona donde la mayoría de la población atendida por el establecimiento educativo se ubica 1=urbano 2=rural
ins_sector	TEXT	Naturaleza administrativa de establecimiento educativo según 1=oficial urbano 2=oficial rural

		3=no oficial
ins_tipo_estab	TEXT	Tipo de establecimiento 1=Oficial urbano 2=Oficial rural 3=no oficial
ins_calendario	TEXT	Calendario del establecimiento
ins_nivel_socio	TEXT	NSE asignado de acuerdo a la clasificación realizada con puntajes promedios

Fuente: Esta investigación.

Tabla 7. Diccionario de Datos de la Dimensión Municipios.

Nombre del Atributo	Tipo de dato	Descripción
muni_id_ente	SERIAL	llave primaria
muni_nombre	TEXT	Nombre del municipio
muni_munexclu	TEXT	identificador si el municipio es ETC
muni_depar	TEXT	nombre del departamento al que pertenece

Fuente: Esta investigación.

Tabla 8. Diccionario de Datos de la Dimensión Prueba.

Nombre del Atributo	Tipo de dato	Descripción
id_prueba	SERIAL	llave primaria
prue_estu_consecutivo	TEXT	identificador del estudiante
prue_sn_lenguaje	TEXT	tiene copia en la prueba de lenguaje
prue_sn_matematicas	TEXT	tiene copia en la prueba de matemáticas
prue_sn_ciencias	TEXT	tiene copia en la prueba de ciencias
prue_sn_competencias	TEXT	tiene copia en la prueba de competencias

Fuente: Esta investigación.

Tabla 9. Diccionario de Datos de la Tabla de Hechos.

Nombre del Atributo	Tipo de dato	Descripción
id_estudiante	INTEGER	llave foránea
muni_id_ente	INTEGER	llave foránea
id_institucion	INTEGER	llave foránea

id_prueba	INTEGER	llave foránea
id_hoja	TEXT	identificador del estudiante
numero_estu	INTEGER	número de estudiantes
puntaje_lenguaje	NUMERIC	puntaje lenguaje
puntaje_matematicas	NUMERIC	puntaje matemáticas
puntaje_ciencias	NUMERIC	puntaje ciencias
puntaje_competencias	NUMERIC	puntaje competencias
numero_lenguaje	INTEGER	no. estudiantes prueba lenguaje
numero_matematicas	INTEGER	no. estudiantes prueba matemática
numero_ciencias	INTEGER	no. estudiantes prueba ciencias
numero_competencias	INTEGER	no. estudiantes prueba competencias
Año	VARCHAR	Año en el que se aplicó la prueba

Fuente: Esta investigación.

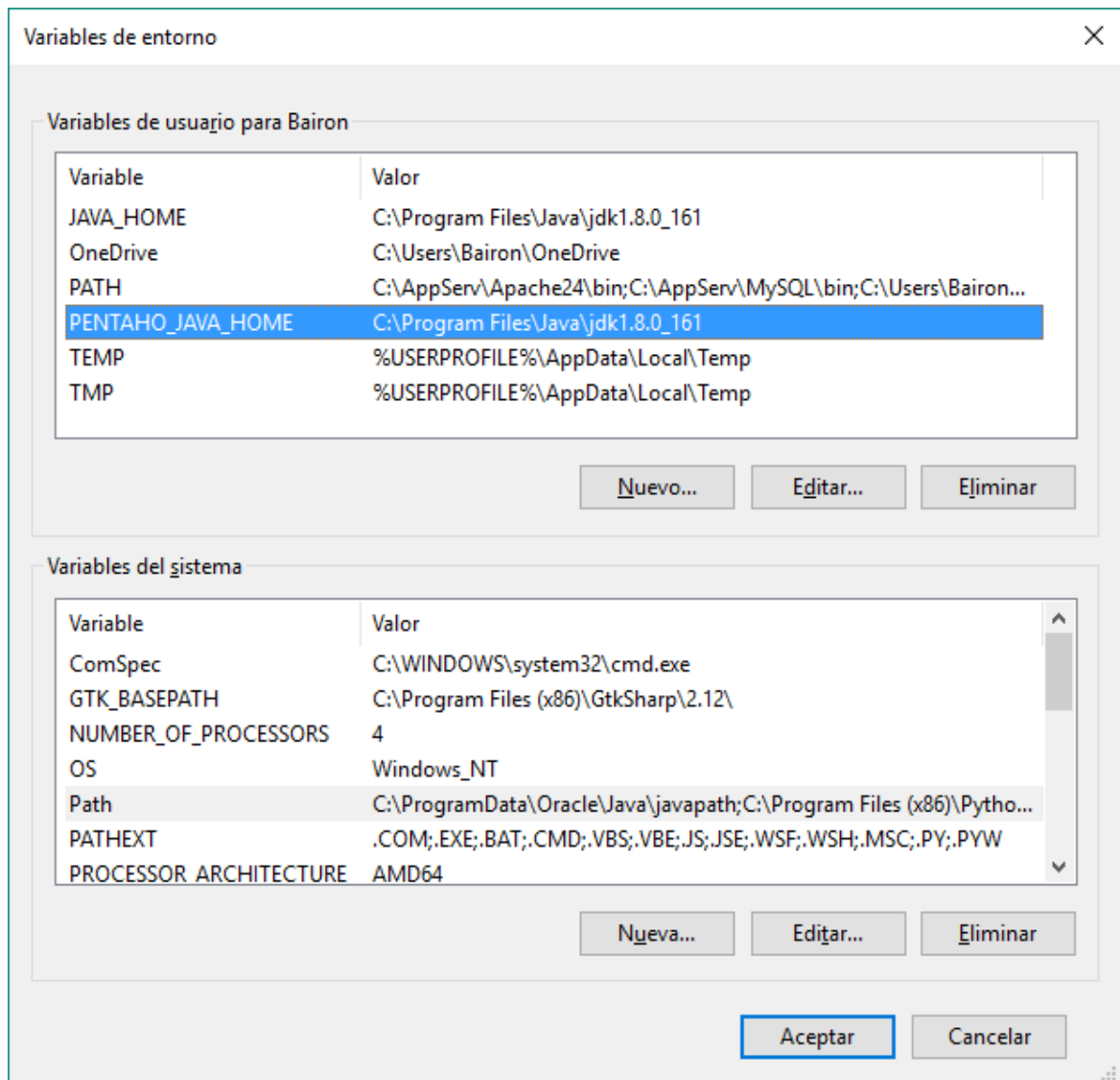
4.4. Preparación de ambiente para el proceso ETL.

Para llevar a cabo el proceso ETL es necesario tener en cuenta los siguientes pasos de configuración e instalación, para no tener inconvenientes a la hora de la ejecución de las aplicaciones necesarias para esta investigación.

A continuación, se describen los pasos necesarios para preparar el ambiente de trabajo:

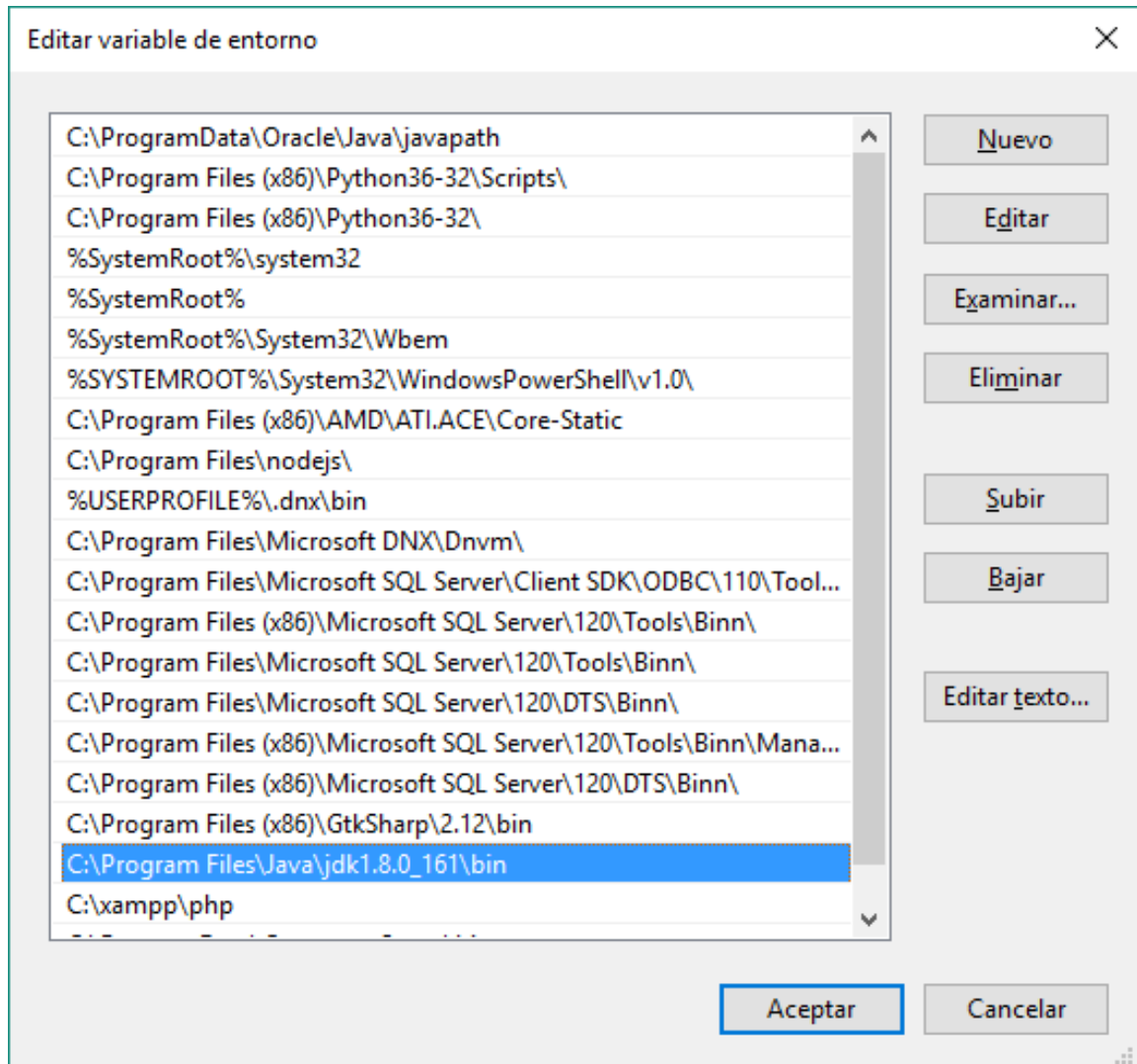
- Versión del Java superior a 7.0.
- Configuración de las Variables de entorno y path, como se describen en la figura 7 y 8.

Figura 7. Configuración del Path.



Fuente: Esta investigación.

Figura 8. Configuración de variables de entorno.



Fuente: Esta investigación.

- Descargar PDI (Pentaho Data Integration) que se ejecuta a través de "Spoon".
- Descargar PD (Pentaho Designer) que se ejecuta a través de "launcher.jar".
- Descargar bi-server de Pentaho que se ejecuta a través de "start-pentaho.bat".
- Descargar Scheme workbench de Mondrian que se ejecuta a través de "workbench.bat".

4.5. Diseño del sistema de extracción, transformación y carga (ETL)

Para implementar el Datamart se usó el motor de bases de datos relacionales PostgreSQL Versión 9.5.1; inicialmente se usaron tres repositorios de datos suministrados por el FTP del ICFES, cada uno con información referente a las pruebas Saber 5° para los años 2014, 2015 y 2016.

A continuación, se describe el proceso de Extracción, Transformación y Carga para construir y poblar el Datamart.

4.5.1. Fase de extracción. El propósito general de esta fase fue construir un repositorio de datos limpio y estandarizado a partir de los datos fuentes, entregados para tal fin se desarrollaron tres guías de trabajo.

4.5.1.1. Integración de Datos. Se creó el repositorio saber_5 usando del repositorio fuente entregado; fue necesario revisar los diccionarios de datos e identificar los atributos de cada una de ellas ya que no existía un 100% de coincidencia entre el documento y las columnas de las tablas. En la Tabla 10 se presenta el número de registros que contenía cada repositorio.

Tabla 10. Número de registros en cada repositorio.

Repositorio	No. de registros
Valores_plausibles_2014_2016	2'162.081
Entidades	95
Establecimientos	20.198
Municipios	1.061
TOTAL, REGISTROS	2'183.435

Fuente: Esta investigación.

En la Tabla 11, 12, 13 y 14 se presentan los atributos de los Valores plausibles, Entidades territoriales certificadas, Establecimientos y Municipios existentes en los repositorios.

Tabla 11. Valores Plausibles.

VARIABLES	2014	2015	2016	DESCRIPCIÓN DE VARIABLES
estu_consecutivo	X	X	X	no está definido en el diccionario de datos
aplicación	X	X	X	no está definido en el diccionario de datos
calendario	X	X	X	Calendario del establecimiento
departamento	X	X	X	Código DANE del departamento al que pertenece el establecimiento educativo
dissenso	X	X	X	Marca de discapacidad cognitiva
enteterr	X	X	X	Código DANE de la entidad a la que pertenece el establecimiento educativo
establecimiento	X	X	X	Código DANE del establecimiento educativo al que pertenece el estudiante
estrato	X	X	X	Estrato al que pertenecen el establecimiento dentro del marco muestral
grado	X	X	X	Grado del estudiante
grupo	X	X	X	Grupo/salón al que pertenecen los estudiantes
id_hoja	X	X	X	Código de identificación de hoja de respuestas
jornada	X	X	X	Código de la sede jornada al que pertenece el estudiante
leng_copietas	X	X	X	Indicador de copia en lenguaje
leng_score1	X	X	X	Score1-valor plausible 1 de lenguaje
leng_score2	X	X	X	Score2-valor plausible 2 de lenguaje
leng_score3	X	X	X	Score3-valor plausible 3 de lenguaje
leng_score4	X	X	X	Score4-valor plausible 4 de lenguaje
leng_score5	X	X	X	Score5-valor plausible 5 de lenguaje
leng_weight	X	X	X	Peso muestral del lenguaje
mate_copietas	X	X	X	Indicador de copia en matemáticas
mate_score1	X	X	X	Score1-valor plausible 1 de matemáticas
mate_score2	X	X	X	Score2-valor plausible 2 de matemáticas
mate_score3	X	X	X	Score3-valor plausible 3 de matemáticas
mate_score4	X	X	X	Score4-valor plausible 4 de matemáticas
mate_score5	X	X	X	Score5-valor plausible 5 de matemáticas
mate_weight	X	X	X	Peso muestral de matemáticas
cien_copietas	X			Indicador de copia en ciencias
cien_score1	X			Score1-valor plausible 1 de ciencias

cien_score2	X			Score2-valor plausible 2 de ciencias
cien_score3	X			Score3-valor plausible 3 de ciencias
cien_score4	X			Score4-valor plausible 4 de ciencias
cien_score5	X			Score5-valor plausible 5 de ciencias
cien_weight	X			Peso muestral de ciencias
comp_copietas		X	X	Indicador de copia en competencias
comp_score1		X	X	Score1-valor plausible 1 de competencias
comp_score2		X	X	Score2-valor plausible 2 de competencias
comp_score3		X	X	Score3-valor plausible 3 de competencias
comp_score4		X	X	Score4-valor plausible 4 de competencias
comp_score5		X	X	Score5-valor plausible 5 de competencias
comp_weight		X	X	Peso muestral de competencias
modeloedu	X	X	X	no está definido en el diccionario de datos
municipio	X	X	X	Código DANE al que pertenece el establecimiento educativo
n	X	X	X	Matricula del grupo al que pertenecen
nivel	X	X	X	Nivel socioeconómico del establecimiento
sector	X	X	X	Sector del establecimiento
sexo	X	X	X	Sexo del estudiante
zona	X	X	X	Zona de la jornada
zonastab	X	X	X	Zona del establecimiento

Fuente: Esta investigación.

Tabla 12. Entidades.

VARIABLES	2014	2015	2016	DESCRIPCIÓN DE VARIABLES
id_ente	X	X	X	Identificador del ente territorial
nombre	X	X	X	Nombre del ente territorial
munexclu	X	X	X	Nombre de los municipios certificados excluidos dentro del ente territorial
tipo	X	X	X	Tipo de ente territorial 1=ETC, 2=DPTO,3= MPIO

Fuente: Esta investigación.

Tabla 13. Establecimientos.

VARIABLES	2014	2015	2016	DESCRIPCIÓN DE VARIABLES
cod_dane	X	X	X	Código DANE del establecimiento educativo
id_municipio	X	X	X	Código DANE del municipio al que pertenece el establecimiento educativo
id_ente	X	X	X	id de la entidad territorial
nombre	X	X	X	Nombre del establecimiento educativo reportado en el DUE
zona	X	X	X	Zona donde la mayoría de la población atendida por el establecimiento educativo se ubica 1=urbano 2=rural
sector	X	X	X	Naturaleza administrativa de establecimiento educativo según 1=oficial urbano 2=oficial rural 3=no oficial
tipo_estab	X	X	X	Tipo de establecimiento 1=Oficial urbano 2=Oficial rural 3=no oficial
calendario	X	X	X	Calendario del establecimiento
nivel_socio	X	X	X	NSE asignado de acuerdo a la clasificación realizada con puntajes por medios

Fuente: Esta investigación.

Tabla 14. Municipios.

VARIABLES	2014	2015	2016	DESCRIPCIÓN DE VARIABLES
id_ente	X	X	X	Identificador del ente territorial
nombre	X	X	X	Nombre del ente territorial
munexclu	X	X	X	Nombre de los municipios certificados excluidos dentro del ente territorial
tipo	X	X	X	Tipo del ente territorial posibles valores 3

Fuente: Esta investigación.

4.5.1.2. Análisis de la calidad de datos

Análisis calidad de datos de la bodega de datos. Se creó la tabla saber5_limpio a partir de la tabla saber5_2014_2016 con el propósito de realizar un análisis inicial de la calidad de los datos e identificar el número de registros no nulos, nulos, distintos valores y valores distintos. Este repositorio cuenta con 65 atributos y 2'183.435 registros.

En las siguientes tablas se muestran los resultados del análisis.

Tabla 15. Análisis de la calidad de Valores plausibles.

No.	Atributos	# No nulos	# Nulos	# Valores distintos	Valores distintos
1	id_hoja	2162081	0	2162081	V201450000001
					V201450000002
					V201450000003
					V201450000004
					V201450000005
2	grupo	2162081	0	14	"01"
					"02"
					"03"
					"04"
					"05"
3	n	2162081	0	389	"1"
					"10"
					"100"
					"101"
4	estrato	2162081	0	29	"102"
					"0000011"
					"0000021"
					"0000092"
					"0000099"
5	aplicación	2162081	0	1	"00000MA"
6	grado	2162081	0	1	"NO_DEFINIDO"
7	jornada	2162081	0	43733	"5"
					"178188"
					"178230"
					"178263"
					"178290"
8	establecimiento	2162081	0	18531	"178293"
					"034100105073"
					"105001000001"
					"105001000043"
					"105001000132"
9	enteterr	2162081	0	99	"105001000141"
					"01"
					"02"

					"03"
					"04"
					"05"
10	departamento	2162081	0	35	"05"
					"08"
					"11"
					"13"
					"15"
11	municipio	2162081	0	1127	"05001"
					"05002"
					"05004"
					"05021"
					"05030"
12	sexo	2162081	0	3	"FEMENINO"
					"MASCULINO"
					"NO_ESPECIFICA"
13	sector	2162081	0	2	"NO_OFICIAL"
					"OFICIAL"
14	zona	2162081	0	2	"RURAL"
					"URBANA"
15	zonastab	2162081	0	2	"RURAL"
					"URBANA"
16	calendario	2162081	0	2	"A"
					"B"
17	nivel	2162081	0	5	"1"
					"2"
					"3"
					"4"
					"5"
18	modeloedu	2162081	0	1	"SIN_DATO"
19	disenso	2162081	0	2	"N"
					"S"
20	leng_copietas	1436436	725645	2	"CON_COPIA"
					"NO_COPIA"
21	leng_weight	1436436	725645	2358	"1"
					"1,09091"
					"1,1"
					"1,11"
					"1,11765"
22	leng_score1	1436436	725645	412999	"-0,00001"
					"-0,00002"
					"-0,00003"
					"-0,00004"
					"-0,00005"
23	leng_score2	1436436	725645	413240	"-0,00001"
					"-0,00002"
					"-0,00003"
					"-0,00005"
24	leng_score3	1436436	725645	413583	"-0,00006"
					"-0,00014"

					"-0,00015"
					"-0,00016"
					"-0,00017"
					"-0,00018"
25	leng_score4	1436436	725645	413559	"-0,0009"
					"-0,00091"
					"-0,00092"
					"-0,00093"
					"-0,00094"
26	leng_score5	1436436	725645	413308	"-0,00034"
					"-0,00035"
					"-0,00036"
					"-0,00037"
					"-0,00038"
27	mate_copietas	1433946	728135	2	"CON_COPIA"
					"NO_COPIA"
					"NULL"
					"1"
28	mate_weight	1433946	728135	2387	"1,06061"
					"1,1"
					"1,11765"
					"1,125"
29	mate_score1	1433946	728135	43,588	"-0,00001"
					"-0,00004"
					"-0,00005"
					"-0,00006"
					"-0,00009"
30	mate_score2	1433946	728135	430721	"-0,00041"
					"-0,00042"
					"-0,00043"
					"-0,00044"
					"-0,00045"
31	mate_score3	1433946	728135	430333	"-0,00001"
					"-0,00002"
					"-0,00003"
					"-0,00004"
					"-0,00005"
32	mate_score4	1433946	728135	430710	"-0,00001"
					"-0,00002"
					"-0,00003"
					"-0,00004"
					"-0,00005"
33	mate_score5	1433946	728135	430798	"-0,00023"
					"-0,00024"
					"-0,00025"
					"-0,00026"
					"-0,00027"
34	comp_copietas	457333	1704748	2	"CON_COPIA"
					"NO_COPIA"
					"NULL"

35	comp_weight	457333	1704748	1577	"1"
					"1,07692"
					"1,09375"
					"1,1"
36	comp_score1	457333	1704748	268532	"1,125"
					"-0,0005"
					"-0,00052"
					"-0,00053"
37	comp_score2	457333	1704748	268662	"-0,00054"
					"-0,00055"
					"-0,0006"
					"-0,00062"
38	comp_score3	457333	1704748	268586	"-0,00064"
					"-0,00065"
					"-0,00066"
					"-0,0008"
39	comp_score4	457333	1704748	268673	"-0,00081"
					"-0,00082"
					"-0,00083"
					"-0,00084"
40	comp_score5	457334	1704749	268909	"-0,00071"
					"-0,00072"
					"-0,00073"
					"-0,00075"
34	cien_copietas	967366	1194715	2	"-0,00076"
					"-0,00001"
					"-0,00002"
					"-0,00003"
35	cien_weight	967366	1194715	1707	"-0,00004"
					"-0,00005"
					"CON_COPIA"
					"NO_COPIA"
36	cien_score1	967366	1194715	372457	"NULL"
					"1"
					"1,11"
					"1,12"
37	cien_score2	967366	1194715	372173	"1,14"
					"1,17"
					"-0,00001"
					"-0,00002"
38	cien_score3	967366	1194715	372114	"-0,00003"
					"-0,00004"
					"-0,00005"
					"-0,0002"
37	cien_score2	967366	1194715	372173	"-0,00062"
					"-0,00021"
					"-0,00022"
					"-0,00023"
38	cien_score3	967366	1194715	372114	"-0,00001"
					"-0,00002"

					"-0,00003"
					"-0,00004"
					"-0,00005"
39	cien_score4	967366	1194715	372138	"-0,001"
					"-0,00101"
					"-0,00103"
					"-0,00104"
					"-0,00105"
40	cien_score5	967366	1194715	373002	"-0,0019"
					"-0,00191"
					"-0,00192"
					"-0,00193"
					"-0,00194"

Fuente: Esta investigación.

Tabla 16. Análisis de la Calidad de Municipios.

No.	Atributos	# No nulos	# Nulos	# Valores distintos	Valores distintos
1	id_ente	1061	0	1061	"05002"
					"05004"
					"05021"
					"05030"
					"05031"
2	nombre	1061	0	978	ABEJORRAL
					ABREGO
					BELARCAZAR
					BECERRIL
					BOYACA
3	tipo	1061	0	1	SIN_DATO

Fuente: Esta investigación.

Tabla 17. Análisis de la Calidad de Instituciones.

No.	Atributos	# No nulos	# Nulos	# Valores distintos	Valores distintos
1	cod_dane	20198	0	20198	"1050010000001"
					"234100105073"
					"105001000108"
					"105001000167"
					"105001000205"
2	id_municipio	20198	0	1121	"05001"
					"05002"

					"05004"
					"05021"
					"050034"
3	id_ente	20198	0	95	"22"
					"23"
					"24"
					"25"
4	zona	20198	0	2	RURAL
					URBANO
5	Sector	20198	0	2	NO OFICIAL
					OFICIAL
6	tipo_estab	20198	0	3	NO OFICIAL
					OFICIAL RURAL
					OFICIAL URBANO
7	calendario	20198	0	2	A
					B
8	nivel_socio	20198	0	5	1
					2
					3
					4
					9

Fuente: Esta investigación.

Análisis calidad de datos del Datamart, se creó el Datamart con las 4 dimensiones y la tabla de hechos con el propósito de realizar un análisis de la calidad de los datos e identificar que los datos contenidos en este Datamart sean fiables, para ponerlos a disposición de los entes académicos. Este Datamart cuenta con 42 atributos y 41.996 registros.

En las siguientes tablas se muestran los resultados del análisis.

Tabla 18. Análisis de la Calidad de Dimensión prueba.

No.	Atributos	# No nulos	# Nulos	# Valores distintos	Valores distintos
1	id_prueba	13875	0	13875	"40"
					"41"
					"42"
					"43"
					"44"
2	prue_estu_consecutivo	13875	0	13875	"V201450046302"
					"V201450046303"

					"V201450046304"
					"V201450046306"
					"V201450046309"
3	prue_sn_lenguaje	9056	4819	2	"CON_COPIA"
					"NO_COPIA"
4	prue_sn_matematicas	9170	4705	2	"CON_COPIA"
					"NO_COPIA"
5	prue_sn_ciencias	6051	7824	2	"CON_COPIA"
					"NO_COPIA"
6	prue_sn_competencias	3011	10864	2	"CON_COPIA"
					"NO_COPIA"

Fuente: Esta investigación.

Tabla 19. Análisis de la Calidad de Dimensión Municipio.

No.	Atributos	# No nulos	# Nulos	# Valores distintos	Valores distintos
1	muni_id_ente	13	0	13	"52022"
					"52210"
					"522152"
					"52224"
					"52227"
2	muni_nombre	13	0	13	"ALDANA"
					"CONTADERO"
					"CORDOBA"
					"CUASPUD"
					"CUMBAL"
3	muni_munexclu	13	0	2	"ETC"
					"SIN_DATO"
4	muni_depar	13	0	1	"NARINO"

Fuente: Esta investigación.

Tabla 20. Análisis de la Calidad de Dimensión estudiante.

No.	Atributos	# No nulos	# Nulos	# Valores distintos	Valores distintos
1	id_estudiante	13875	0	13875	"50"
					"51"
					"52"
					"53"

					"54"
2	estu_consecutivo	13875	0	13875	"V201450046302"
					"V201450046303"
					"V201450046304"
					"V201450046305"
					"V201450046306"
3	estu_grupo	13875	0	3	"01"
					"02"
					"99"
4	estu_n	13875	0	112	"1"
					"10"
					"100"
					"101"
5	estu_estrato	13875	0	12	99E1209
					99E1219
					99E9999
					99O1109
					99O1209
6	estu_grado	13875	0	1	"5"
7	estusexo	13875	0	3	FEMENINO
					MASCULINO
					NO_ESPECIFICA
8	estu_disenso	13875	0	2	"N"
					"S"

Fuente: Esta investigación.

Tabla 21. Análisis de la Calidad de Dimensión institución.

No.	Atributos	# No nulos	# Nulos	# Valores distintos	Valores distintos
1	id_institucion	358	0	358	"39"
					"40"
					"41"
					"42"
					"43"
2	ins_codigo_dane	358	0	358	"152022000084"
					"152210000261"
					"152215000138"

					"152224000019"
					"152227000028"
3	ins_municipio	358	0	13	52022
					52210
					52224
					52227
4	ins_id_ente	358	0	2	48
					89
5	ins_nombre	358	0	336	CENTRO EDUCATIVO SAN MARTIN
					CENTRO EDUCATIVO CASA FRIA
					CENTRO EDUCATIVO ARVELA
					CENTRO EDUCATIVO CUAS
					CENTRO EDUCATIVO GUAN
6	ins_zona	358	0	2	RURAL
					URBANO
7	ins_sector	358	0	2	NO_OFICIAL
					OFICIAL
8	ins_tipo_estab	358	0	3	NO_OFICIAL
					OFICIAL_RURAL
					OFICIAL_URBANO
9	ins_calendario	358	0	1	A
10	ins-nivel_socio	358	0	5	1
					2
					3
					4
					5

Fuente: Esta investigación.

Tabla 22. Análisis de la Calidad de Tabla de hechos.

No.	Atributos	# No nulos	# Nulos	# Valores distintos	Valores distintos
1	id_estudiante	13875	0	13875	"35"
					"36"
					"37"
					"38"
					"39"
2	muni_id_ente	13875	0	13	"52022"
					"52210"
					"52215"
					"52224"
					"52227"
3	id_institucion	13875	0	324	"150"
					"151"
					"163"
					"167"
					"168"
4	id_prueba	13875	0	13875	"1"
					"55"
					"60"
					"70"
					"100"
5	id_hoja	13875	0	13875	"V201450046302"
6	numero_estu	13875	0	1	1
7	puntaje_lenguaje	9056	4819	151	"1"
					"1,2"
					"1,25"
					"1,33"
					"1,38"
8	puntaje_matematicas	9170	4705	154	"1,46"
					"1,525"
					"1,52577"
					"1,53"
					"1,54"
9	puntaje_ciencias	6051	7824	102	"1,45"
					"1,46"
					"1,47"
					"1,48"
					"1,4878"
10	puntaje_competencias	3011	10864	72	"1,5"

					"1,50104"
					"1,50943"
					"1,5283"
					"1,6"
11	numero_lenguaje	13875	0	2	"0"
					"1"
12	numero_matematicas	13875	0	2	"0"
					"1"
13	numero_ciencias	13875	0	2	"0"
					"1"
14	numero_competencias	13875	0	2	"0"
					"1"
15	Año	13875	0	3	"2014,2015,216"

Fuente: Esta investigación.

4.5.1.3. Limpieza de Datos. Se ejecutaron varios procesos sobre las tablas valores_plausibles_2014_2016, municipios_2014_2016, establecimientos_2014_2016, entidades_2014_2016 con el propósito de estandarizar los valores presentes, rellenar los valores nulos, corregir valores erróneos y en general mejorar las condiciones de calidad del repositorio.

Se crearon los repositorios valores_plausibles_2014_2016_limpio, municipios_2014_2016_limpio, establecimientos_2014_2016_limpio, entidades_2014_2016_limpio a partir de los repositorios mencionados anteriormente.

4.5.2. Fase de transformación. El objetivo de esta fase fue construir un repositorio estandarizado y con datos de calidad que fue usado como fuente de carga del Datamart; se desarrollaron dos guías de trabajo.

4.5.2.1. Transformación de Datos. A partir de los repositorios valores_plausibles_2014_2016_limpio, municipios_2014_2016_limpio, establecimientos_2014_2016_limpio, entidades_2014_2016_limpio se crearon las tablas valores_plausibles_2014_2016_final, municipios_2014_2016_final, establecimientos_2014_2016_final, entidades_2014_2016_final, sobre las cuales se ejecutaron procedimientos para normalizar el valor de los atributos y facilitar el posterior análisis de resultados.

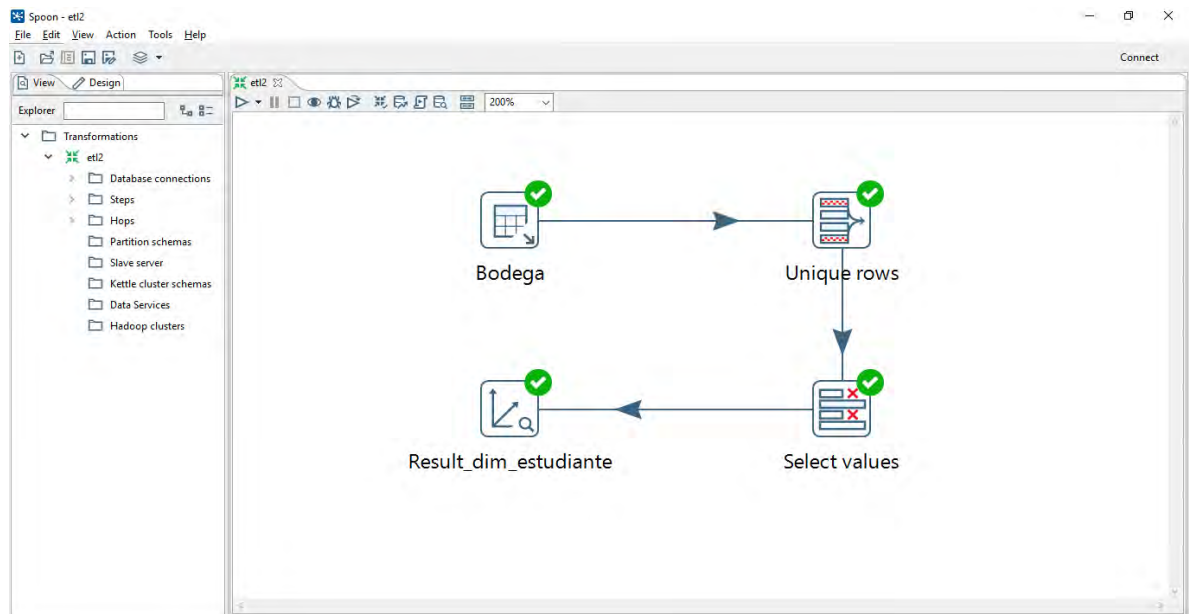
4.5.2.2. Eliminación de Datos. Usando el repositorio valores_plausibles_2014_2016_limpio, municipios_2014_2016_limpio, establecimientos_2014_2016_limpio, entidades_2014_2016_limpio se crearon las tablas valores_plausibles_2014_2016_final, municipios_2014_2016_final, establecimientos_2014_2016_final, entidades_2014_2016_final, sobre la cual se efectuaron rutinas para eliminar atributos que no son requeridos en el Datamart. Finalmente se obtuvo una tabla de 65 atributos.

4.5.3. Fase de carga. Para el proceso de carga de los datos en el Datamart, solo se consideraron aquellos registros relevantes para el fin de esta investigación; a partir de la bodega de datos saber_5 se creó la base de datos final llamado data_mart, el cual contiene las dimensiones y tabla de hechos. La base de datos resultante con la estructura de un Datamart tiene 42 atributos y 41.996 registros. El paso siguiente fue integrar los datos con la herramienta PDI (Pentaho Data Integration) a las cinco tablas del Datamart (4 dimensiones y 1 tabla de hechos).

A continuación se presenta el Proceso ETL automatizado.

- Proceso ETL automatizado para la dimensión estudiante.

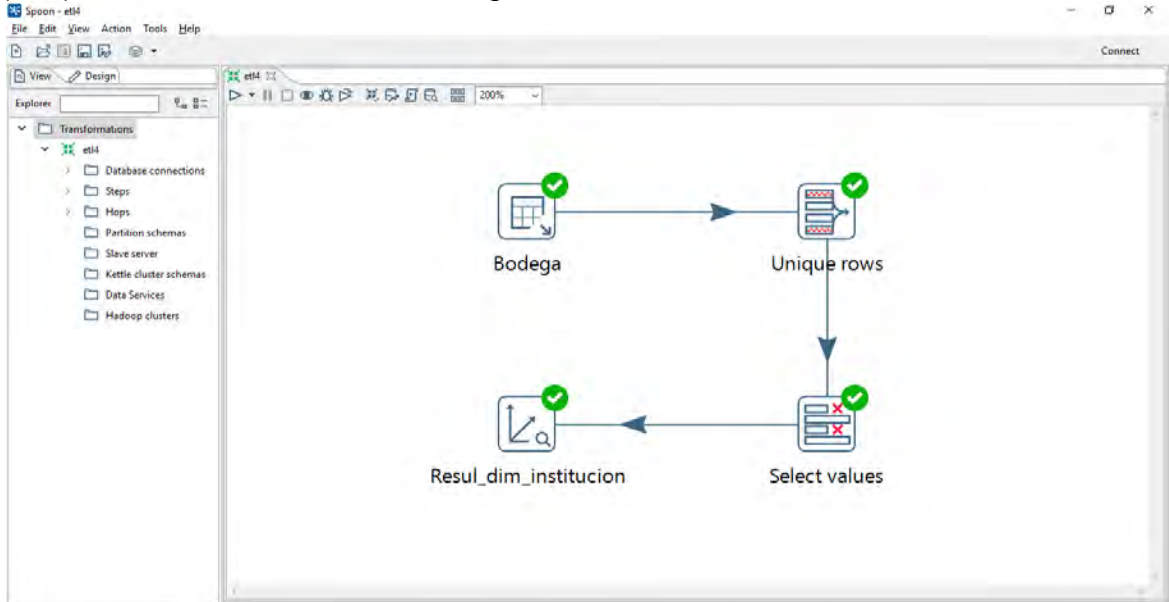
Figura 9. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión estudiante.



Fuente: Esta investigación.

- Proceso ETL automatizado para la dimensión institución.

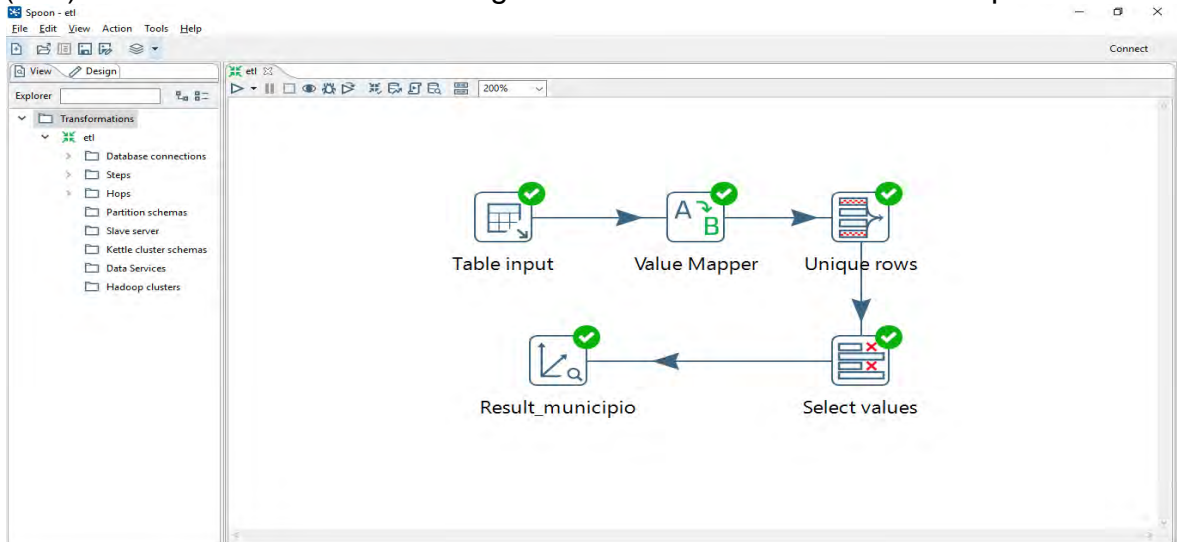
Figura 10. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión institución.



Fuente: Esta investigación.

- Proceso ETL automatizado para la dimensión municipio.

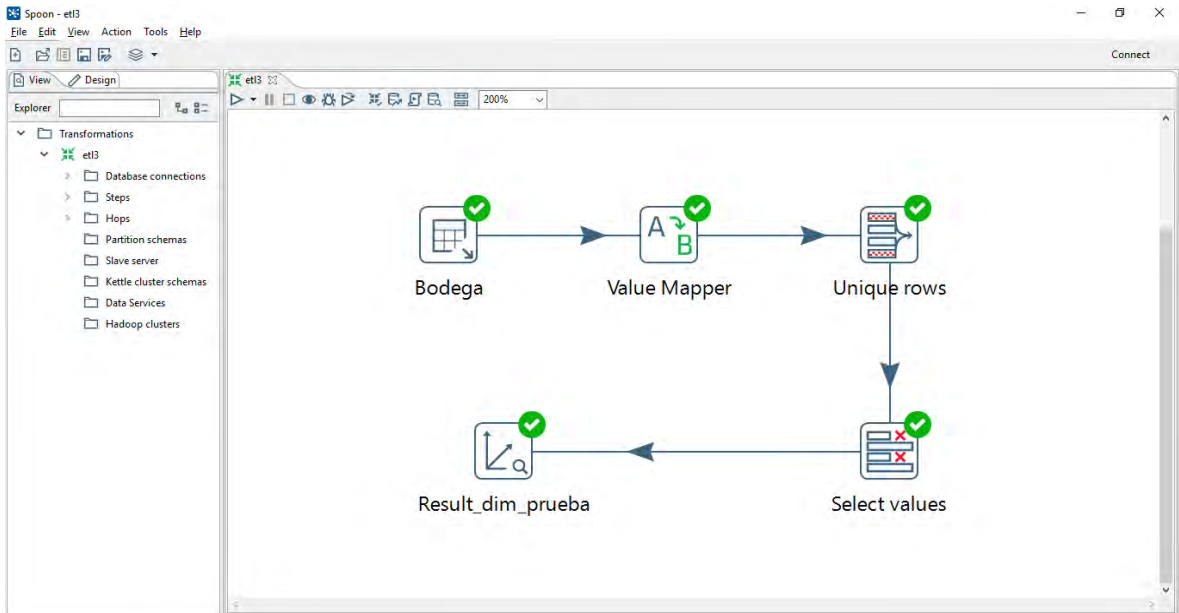
Figura 11. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión municipio.



Fuente: Esta investigación.

- Proceso ETL automatizado para la dimensión prueba.

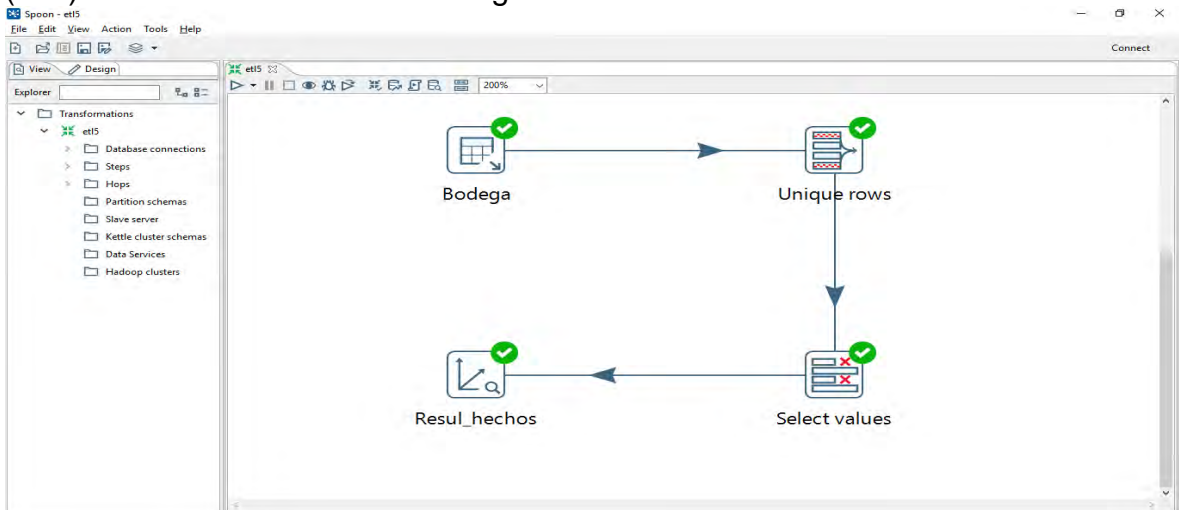
Figura 12. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión prueba.



Fuente: Esta investigación.

- Proceso ETL automatizado para la tabla de hechos.

Figura 13. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la tabla de hechos.



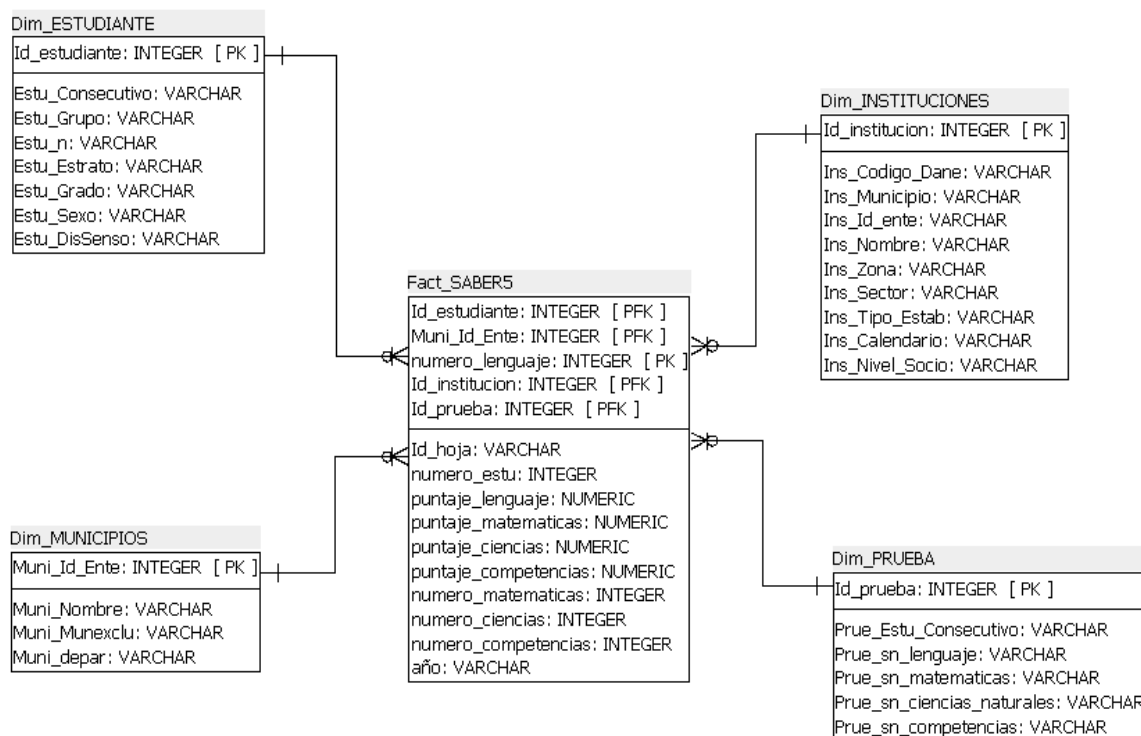
Fuente: Esta investigación.

En las figuras 9 a 13, Se muestra el proceso automatizado ETL, que sirvió para extracción, transformación y carga de datos de la bodega de datos saber_5 al Datamart, demostrando así el uso de los nodos para seleccionar, renombrar, transformar y cargar los atributos más relevantes en esta investigación.

Se tuvo en cuenta, el orden de los nodos para ejecutar el proceso automatizado tal fue el caso de la extracción y carga de las dimensiones y tabla de hechos donde se ejecutó la sentencia SQL para seleccionar los atributos y sus datos de la bodega saber_5, esto dio como resultado los datos relevantes para el proyecto y necesarios para la toma de decisiones, en seguida se realizó la selección y transformación de los atributos con el fin de estandarizar valores iniciales de la bodega, por último se realizó la carga de datos en el Datamart .

4.6. Implementación de la bodega en Postgres

Figura 14. . Implementación del Modelo Multidimensional tipo estrella en PostgreSQL.



Fuente: Esta investigación.

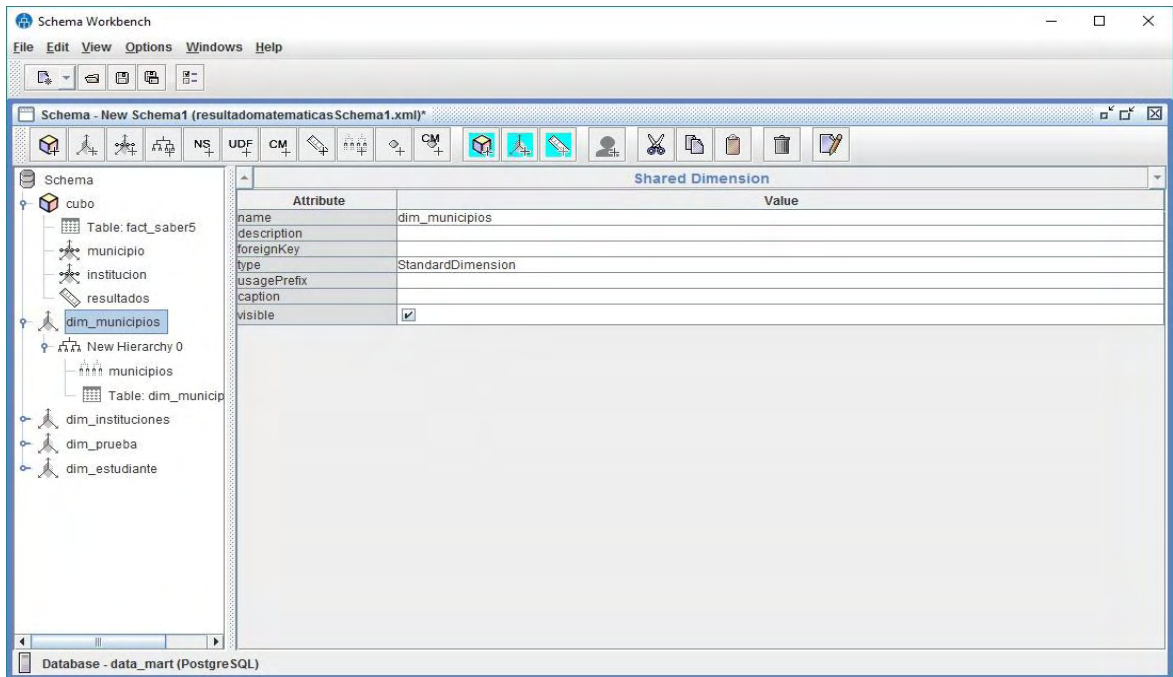
En la figura 14 se presenta el Modelo Multidimensional tipo estrella en PostgreSQL donde podemos observar las 4 dimensiones con sus respectivas llaves primarias y las variables relevantes correspondientes a cada una, además de la tabla de hechos con su respectiva llave primaria, sus llaves foráneas y las variables relevantes.

4.7. Integración del Datamart con Pentaho BI

4.7.1. Cubo de datos. A continuación, en la figura 15 se presenta la configuración del cubo para las consultas OLAP, realizadas en la aplicación Schema Workbench de Mondrian, donde se tuvo en cuenta como primer paso configurar los parámetros de conexión con el Datamart para poder crear el Schema el cual contendrá el cubo y las dimensiones.

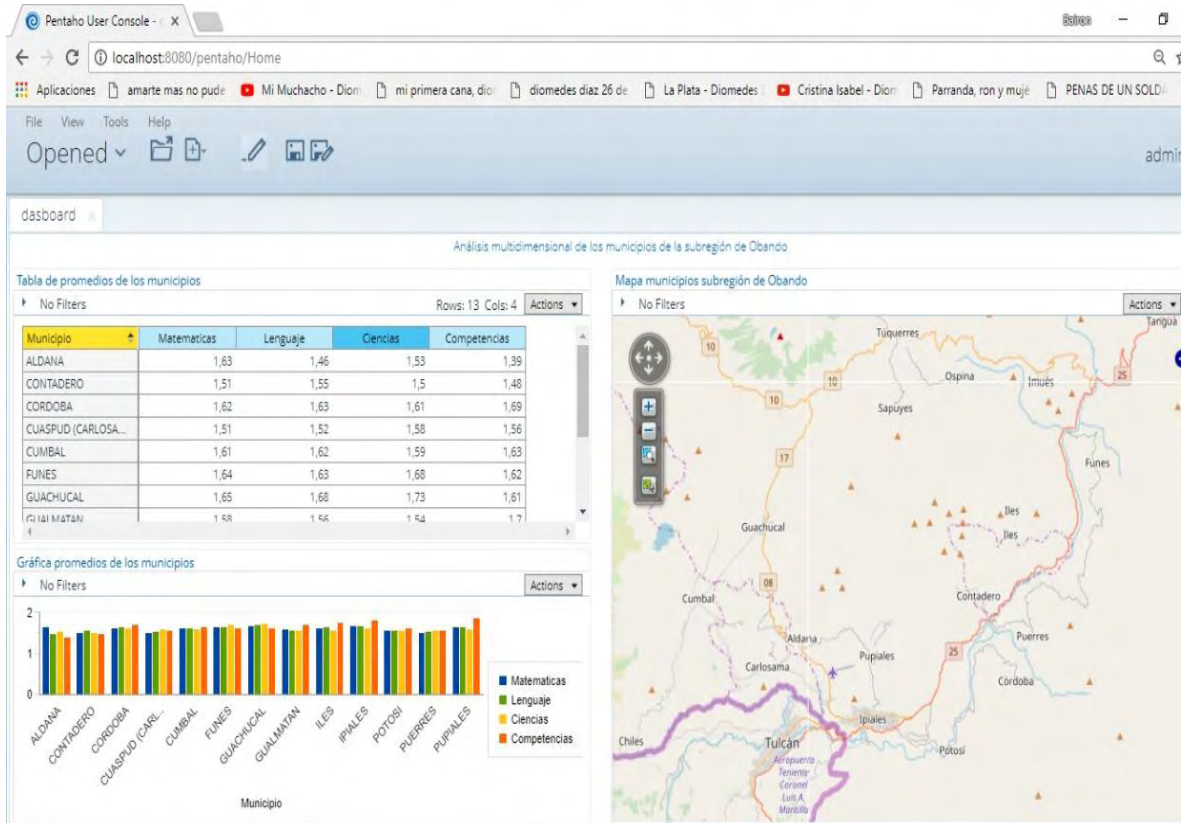
Dentro del Schema creado se tuvo en cuenta el orden de los elementos para que su publicación sea de manera correcta, se creó una a una las dimensiones, configurando su nombre y su tabla de parámetros para acceder al Datamart, por último, se creó el cubo y dentro del mismo se agregó la tabla de hechos y se configuro los parámetros correspondientes para que no existan fallos al momento de publicar el Schema.

Figura 15. Configuración del cubo para las consultas OLAP en Schema Workbench.



Fuente: Esta investigación.

Figura 16. Análisis multidimensional en BI server de Pentaho.



Fuente: Esta investigación.

Para la visualización del dashboard de la figura 16, se tuvo que importar el cubo y se escogió el schema publicado en Workbench de Pentaho donde se consiguió como resultado el dashboard en el cual podemos visualizar el Análisis multidimensional de los resultados de las pruebas saber 5° de las instituciones educativas de los diferentes municipios de la subregión de Obando y su ubicación en el mapa.

4.7.2. Operaciones OLAP. A continuación, en la tabla 23 se presenta las Operaciones OLAP donde podemos observar los promedios de cada municipio en cada una de las áreas.

Tabla 23. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las 4 áreas evaluadas por él ICFES.

Municipio	Promedio Matemáticas	Promedio Lenguaje	Promedio Ciencias	Promedio Competencias
Aldana	1.63	1.46	1.53	1.39
Contadero	1.51	1.55	1.50	1.48
Córdoba	1.62	1.63	1.61	1.69
Cuaspud (Carlosama)	1.51	1.52	1.58	1.56
Cumbal	1.61	1.62	1.59	1.63
Funes	1.64	1.63	1.68	1.62
Guachucal	1.65	1.68	1.73	1.61
Gualmatan	1.58	1.56	1.54	1.70
Iles	1.61	1.63	1.55	1.74
Ipiales	1.67	1.66	1.62	1.81
Potosí	1.55	1.56	1.56	1.60
Puerres	1.51	1.52	1.54	1.55
Pupiales	1.63	1.64	1.59	1.86

Fuente: Esta investigación.

4.8. Especificación y desarrollo de aplicaciones de BI

Para el diseño e implementación de los cubos que representan cada una de las medidas sobre los hechos se utilizó el programa libre bi-server de la suite de Pentaho BI (Pentaho Community Edition).

Después de realizar los procesos requeridos para la configuración, se diseñó cada uno de los cubos usando sentencias SQL y posteriormente se ajustaron las características de visualización. El procedimiento se muestra en las Figuras 17 a 19.

Figura 17. Pantalla de inicio Pentaho.



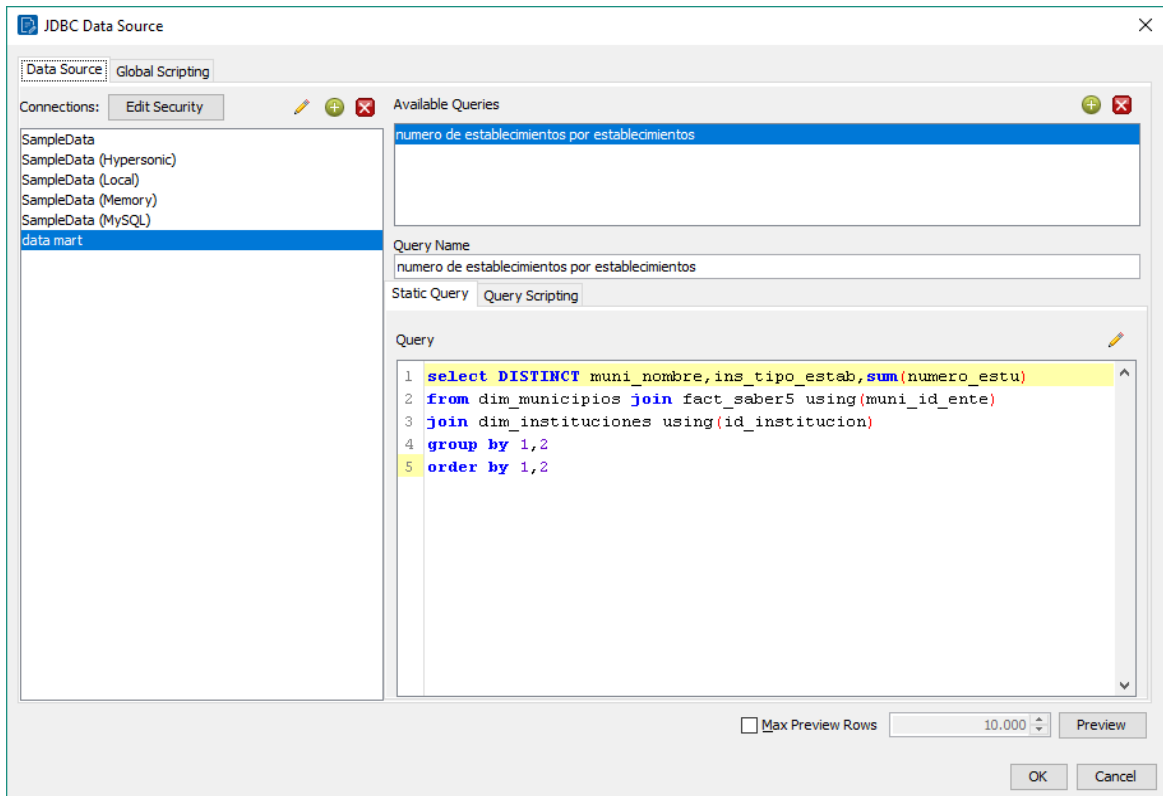
Fuente: Esta investigación.

Figura 18. Creación de reporte en Pentaho.



Fuente: Esta investigación.

Figura 19. Sentencia SQL para diseño de cubos.



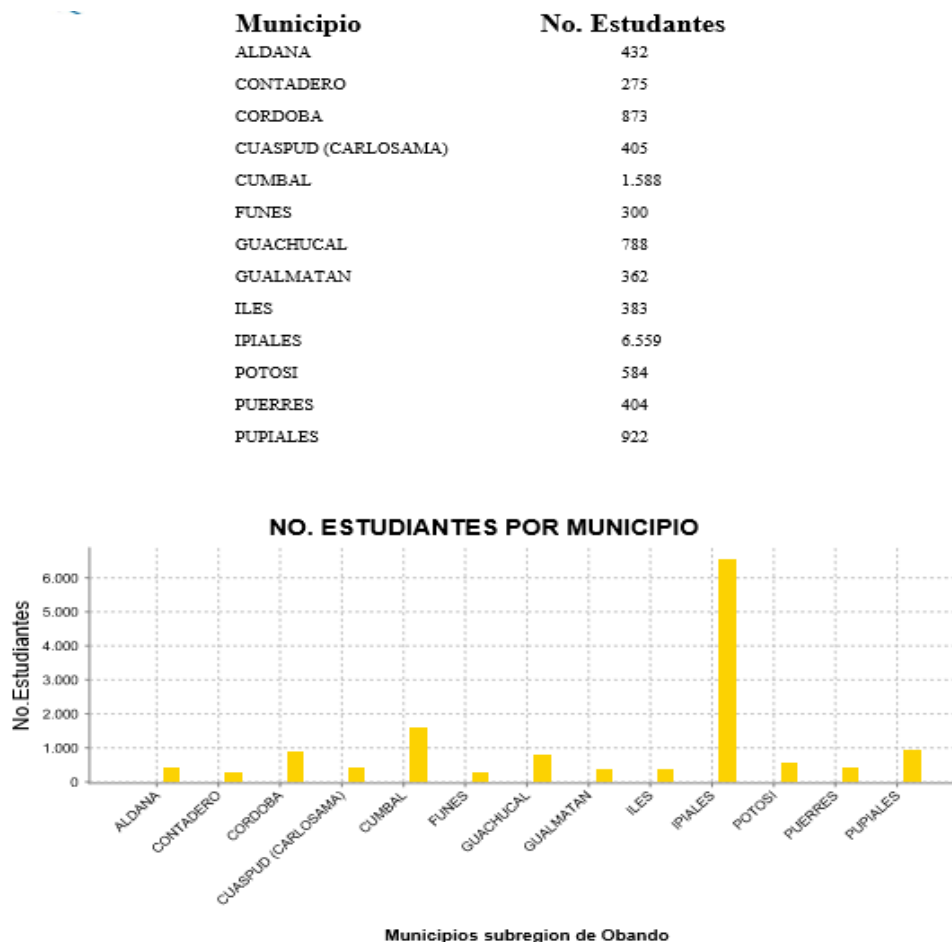
Fuente: Esta investigación.

5. RESULTADOS Y ANÁLISIS

A partir del modelo multidimensional se construyó un cubo para representar cada una de las medidas sobre los hechos. Se realizaron las siguientes consultas:

1. ¿No. estudiantes por municipio que presentaron las pruebas saber 5° entre 2014-2016?

Figura 20. No. estudiantes por municipio que presentaron las pruebas saber 5° entre 2014-2016.

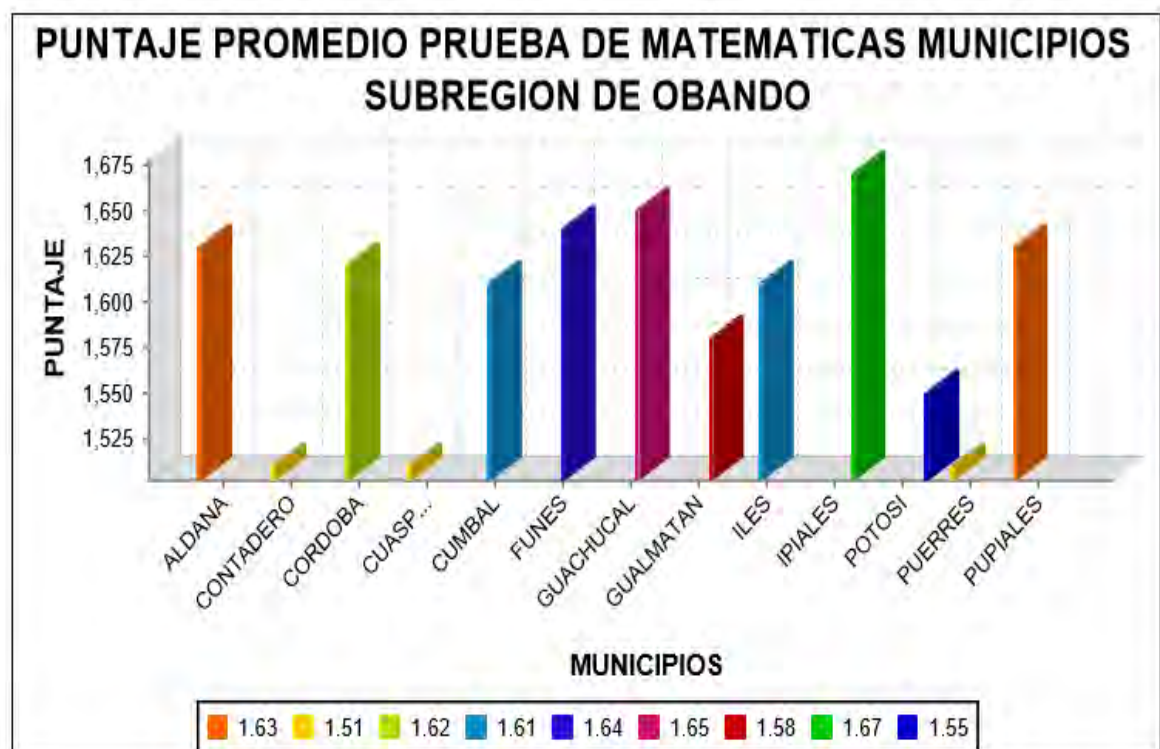


Fuente: Esta investigación.

En figura 20 podemos observar que la subregión de Obando, el municipio de Ipiales tiene el mayor número de estudiantes que presento las pruebas saber 5°, con un porcentaje de 47.27% que representa 6559 estudiantes en los periodos del 2014 al 2016 y el municipio con menor cantidad de estudiantes es el municipio del Contadero con un porcentaje del 1.98% que representa 275 estudiantes en los periodos anteriormente mencionados.

2. ¿Puntaje promedio de los resultados de matemáticas clasificados por Municipios Subregión de Obando?

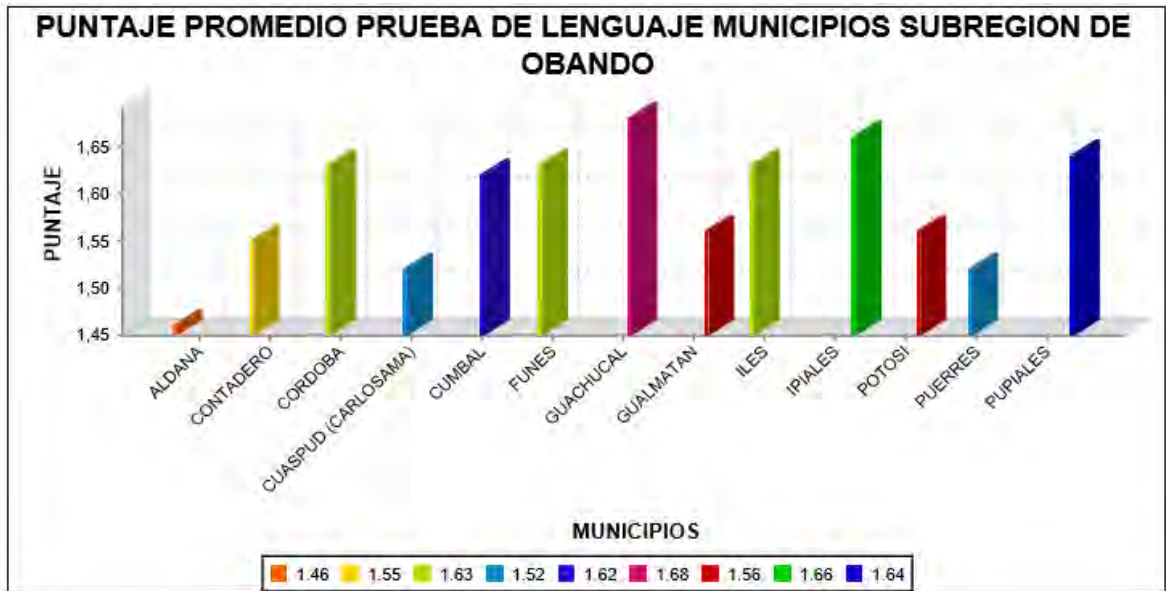
Figura 21. Puntaje promedio de los resultados de matemáticas clasificados por Municipios Subregión de Obando.



Fuente: Esta investigación.

3. ¿Puntaje promedio de los resultados de Lenguaje clasificados por Municipios Subregión de Obando?

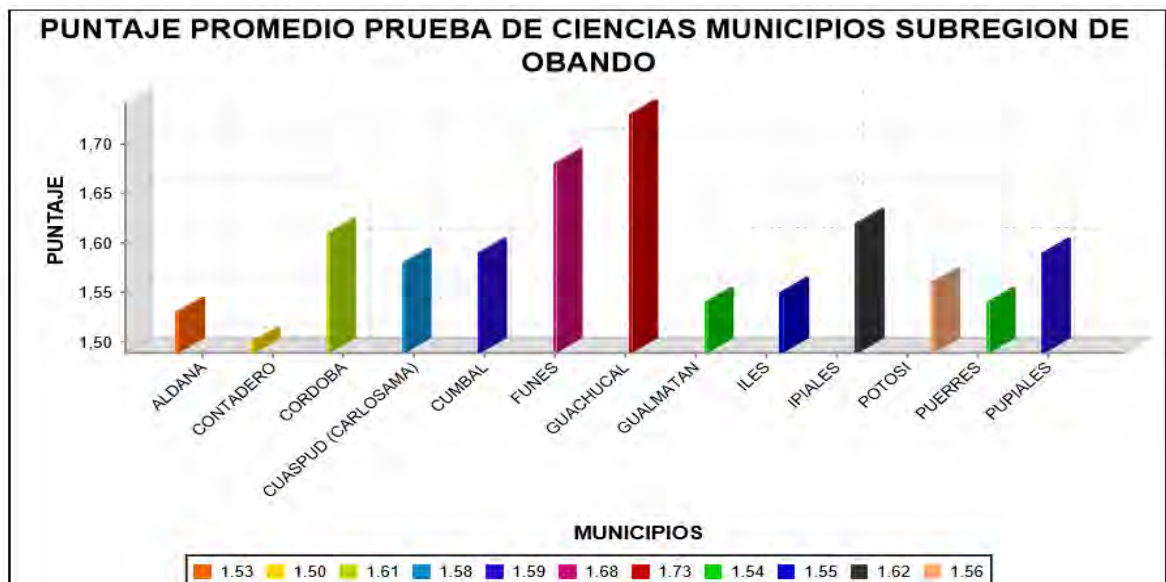
Figura 22. Puntaje promedio de los resultados de Lenguaje clasificados por Municipios Subregión de Obando.



Fuente: Esta investigación.

4. ¿Puntaje promedio de los resultados de Ciencias clasificados por Municipios Subregión de Obando?

Figura 23. Puntaje promedio de los resultados de Ciencias clasificados por Municipios Subregión de Obando.



Fuente: Esta investigación.

5. ¿Puntaje promedio de los resultados de competencias clasificados por Municipios Subregión de Obando?

Figura 24. Puntaje promedio de los resultados de competencias clasificados por Municipios Subregión de Obando.



Fuente: Esta investigación.

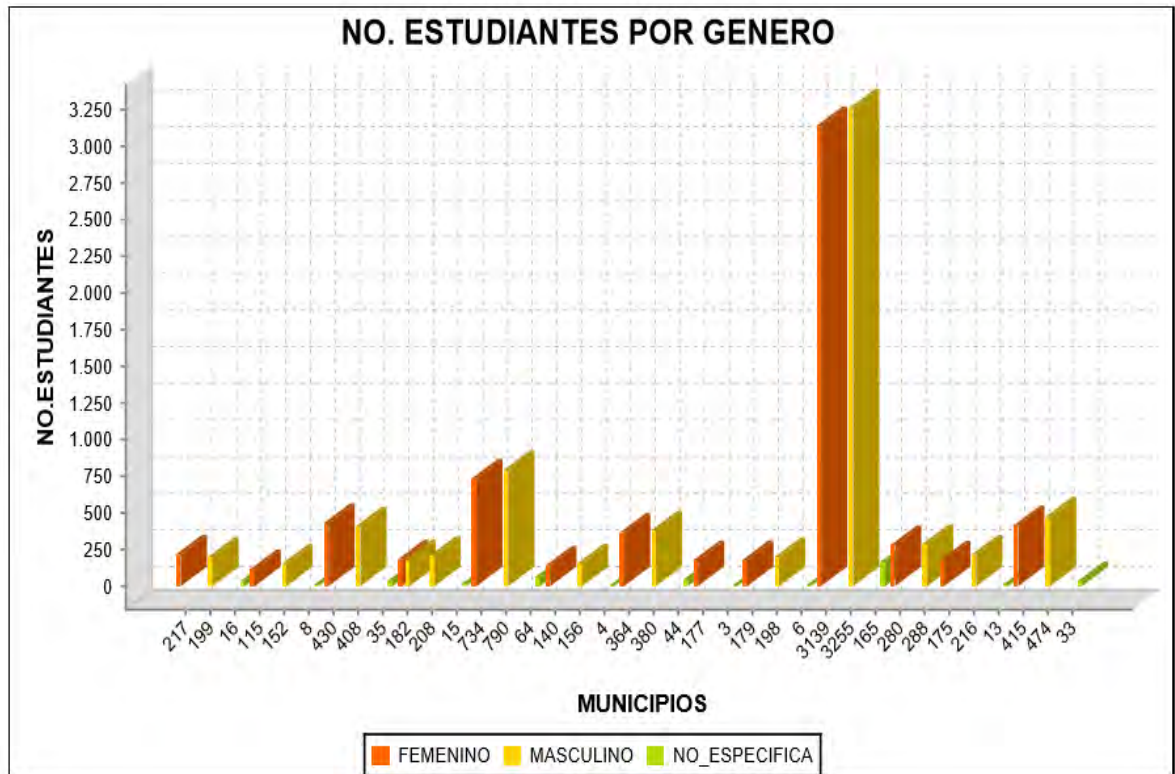
En las figuras 21, 22, 23 y 24 podemos observar el comportamiento de los estudiantes que presentaron las pruebas saber 5° en los años 2014 al 2016 en las áreas de matemáticas, lenguaje, competencias y ciencias, clasificados por municipios de la subregión de Obando, obteniendo los siguientes resultados:

Teniendo en cuenta los resultados generales de cada municipio de la subregión de Obando para el Área de matemáticas, Ipiales tuvo uno de los mejores promedios con 1,67 y Puerres, Contadero y Cuaspud (Carlosama) tienen los promedios más bajos con 1,51. Para el Área de Lenguaje, Guachucal tiene el mejor promedio con 1,68 y Aldana el más bajo con 1,46. Para el Área de Ciencias podemos observar que Guachucal tiene el mejor promedio con 1,73 y Contadero el más bajo con 1,50. Finalmente el área de Competencias tiene a Pupiales con el promedio más alto con 1,86 y Aldana con uno de los más bajos con 1,39.

Para lo anterior podemos concluir que es necesario realizar estrategias de mejoramiento en la calidad académica en todos los municipios de la subregión de Obando, para fortalecer el aprendizaje de los estudiantes de básica primaria ya que los promedios están en un rango bajo con respecto a los resultados a nivel nacional, los municipios que necesitan estas estrategias de manera inaplazable son Contadero y Aldana específicamente.

6. ¿No. Estudiantes que presentaron las pruebas saber 5° clasificados por género?

Figura 25. No. Estudiantes que presentaron las pruebas saber 5° clasificados por género.

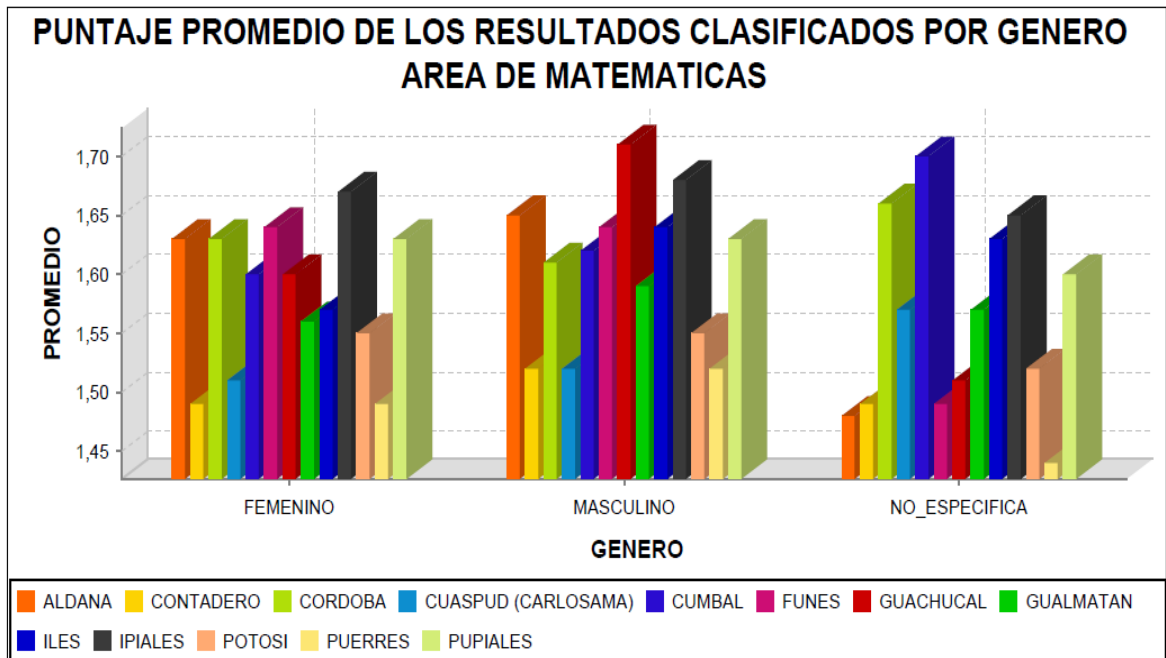


Fuente: Esta investigación.

En la figura 25 podemos observar que hay comportamiento equilibrado entre los estudiantes que presentaron las pruebas saber 5° en los años 2014 al 2016 en los géneros 'Masculino' y 'Femenino', cabe resaltar que la variable 'No_Especific' tiene un comportamiento mínimo frente a las variables mencionadas anteriormente.

7. ¿Puntaje promedio de los resultados por municipio clasificados por género en el área de matemáticas?

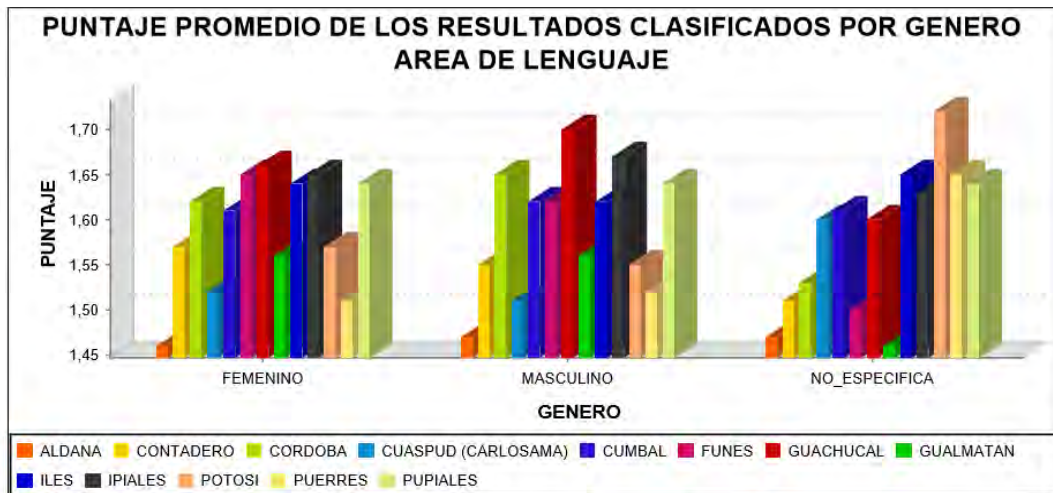
Figura 26. Puntaje promedio de los resultados por municipio clasificados por género en el área de matemáticas.



Fuente: Esta investigación.

8. ¿Puntaje promedio de los resultados por municipio clasificados por género en el área de Lenguaje?

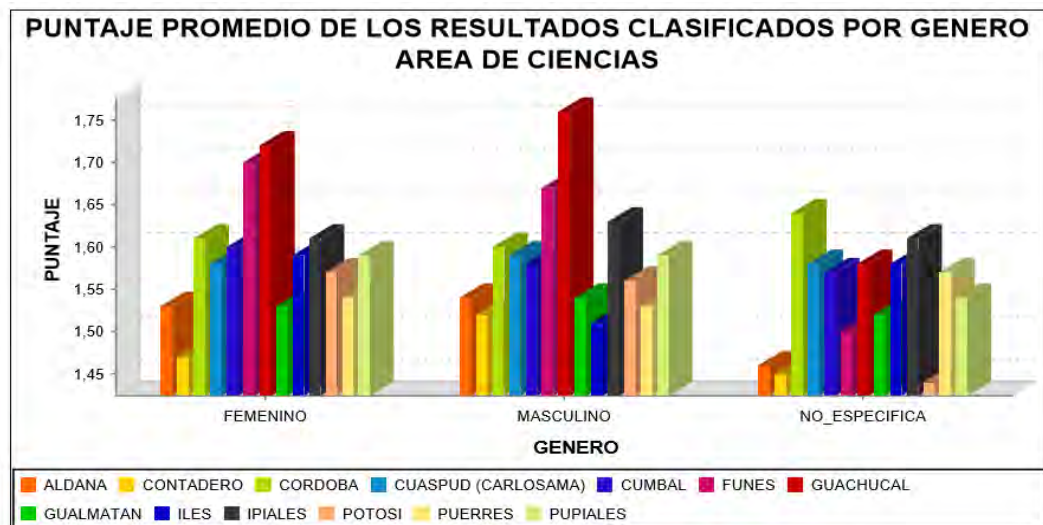
Figura 27. Puntaje promedio de los resultados por municipio clasificados por género en el área de Ciencias.



Fuente: Esta investigación.

9. ¿Puntaje promedio de los resultados por municipio clasificados por género en el área de Ciencias?

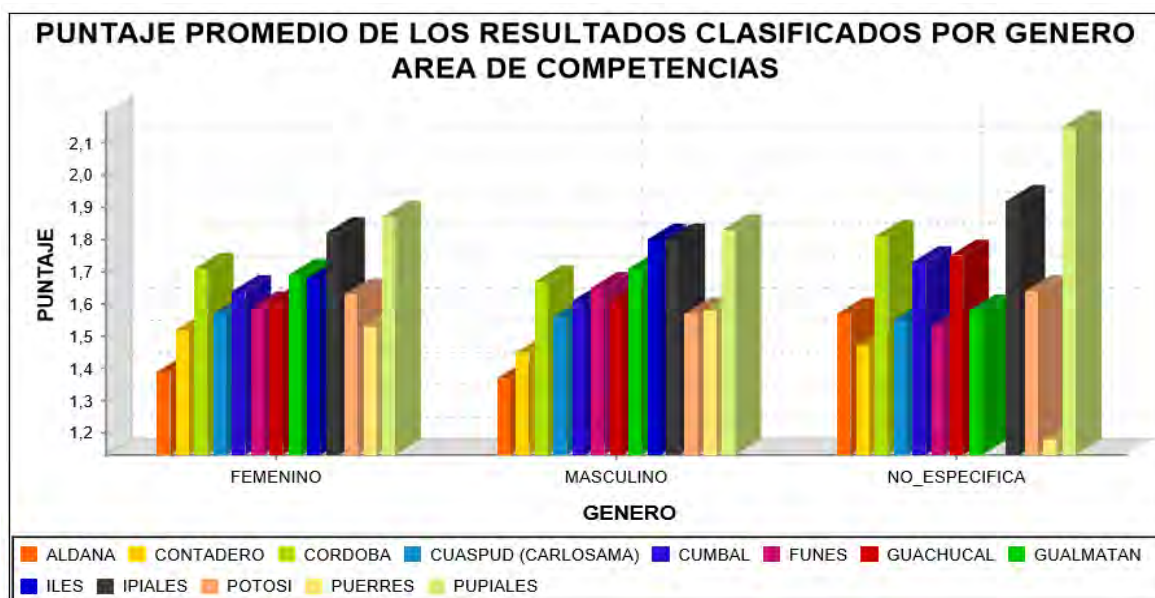
Figura 28. Puntaje promedio de los resultados por municipio clasificados por género en el área de Ciencias.



Fuente: Esta investigación.

10. ¿Puntaje promedio de los resultados por municipio clasificados por género en el área de Competencias?

Figura 29. Puntaje promedio de los resultados por municipio clasificados por género en el área de Competencias.



Fuente: Esta investigación.

En las figuras 26, 27, 28 y 29 podemos observar el comportamiento de los estudiantes que presentaron las pruebas saber 5° en los años 2014 al 2016 en las áreas de matemáticas, lenguaje, competencias y ciencias clasificados por género en cada municipio de la subregión de Obando, obteniendo los siguientes resultados:

Tabla 24. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las áreas, clasificado por géneros.

Áreas	Género			
	Femenino	Masculino	No_Especifica	
Matemáticas	Mayor	Ipiales	Guachucal	Cumbal
	Menor	Contadero Puerres	Contadero Carlosama Puerres	Puerres
Lenguaje	Mayor	Guachucal	Guachucal	Potosí
	Menor	Aldana	Aldana	Gualmatan

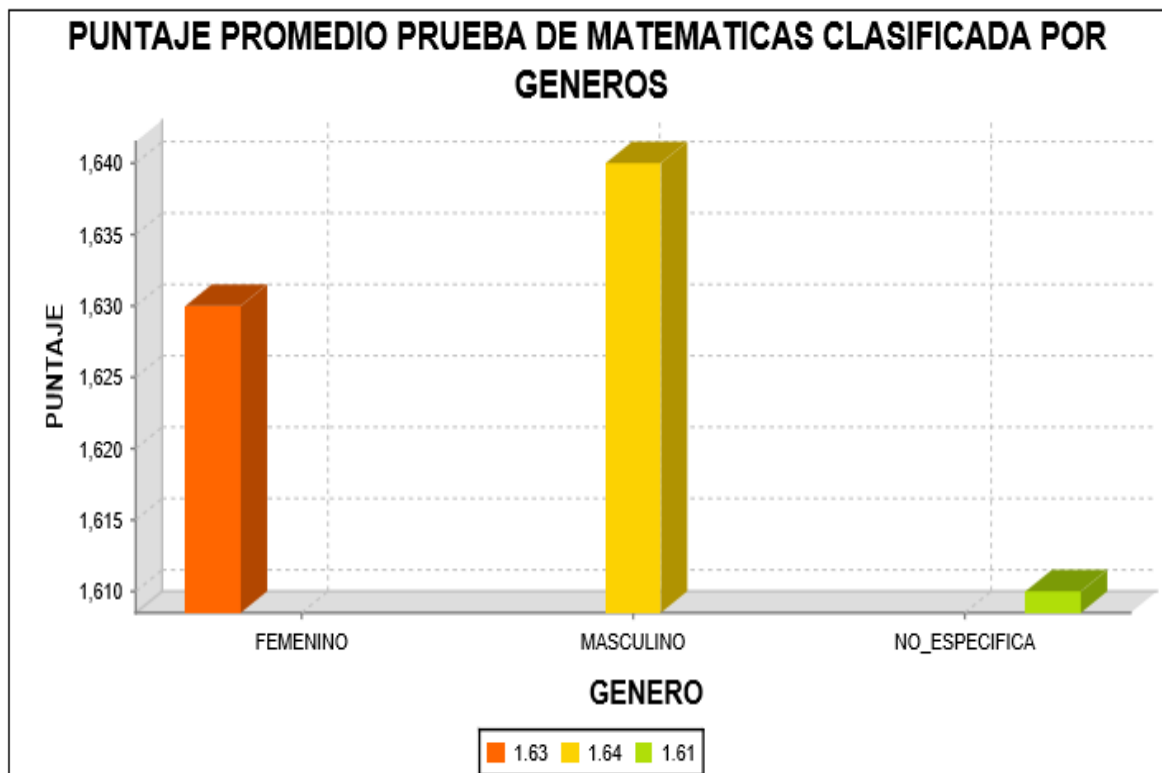
Ciencias	Mayor	Guachucal	Guachucal	Córdoba
	Menor	Contadero	Iles	Potosí
Competencias	Mayor	Pupiales	Pupiales	Pupiales
	Menor	Aldana	Aldana	Puerres

Fuente: Esta investigación.

En la tabla 24 podemos observar que el municipio de Guachucal tuvo los mejores promedios en comparación a los diferentes municipios de la subregión de Obando, otro factor importante tiene que ver con los municipios de Puerres y Aldana los cuales obtuvieron el desempeño más bajo en el área de matemáticas y Lenguaje respectivamente, Por su parte Pupiales tuvo un gran desempeño en el área de competencias.

11. ¿Puntaje promedio de la prueba de Matemáticas clasificados por géneros?

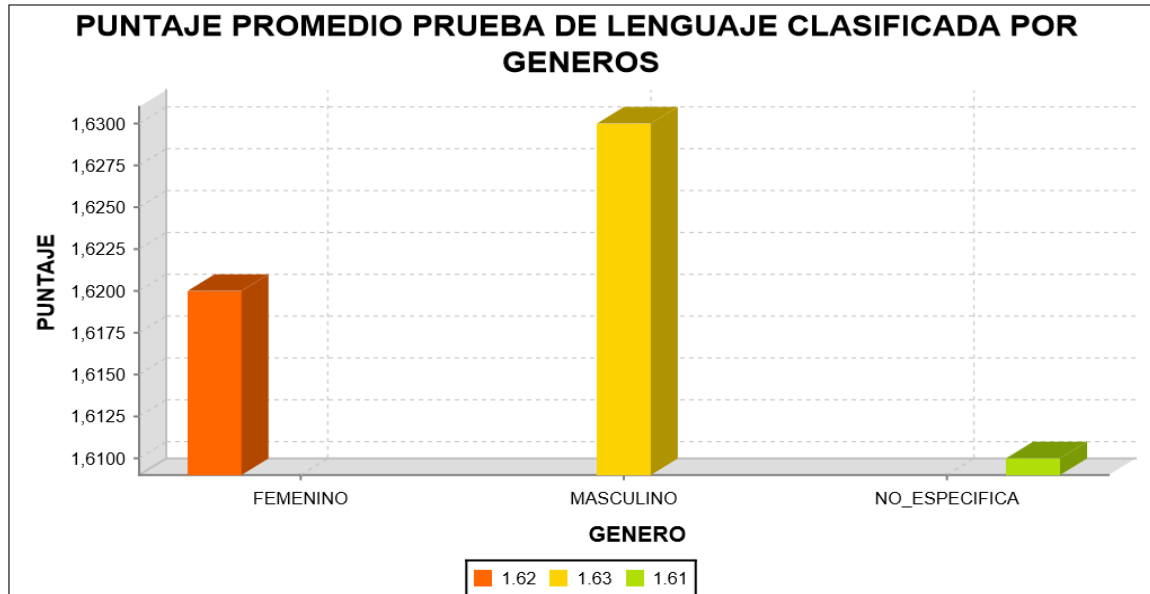
Figura 30. Puntaje promedio de la prueba de Matemáticas, clasificados por géneros.



Fuente: Esta investigación.

12. ¿Puntaje promedio de la prueba de Lenguaje clasificados por géneros?

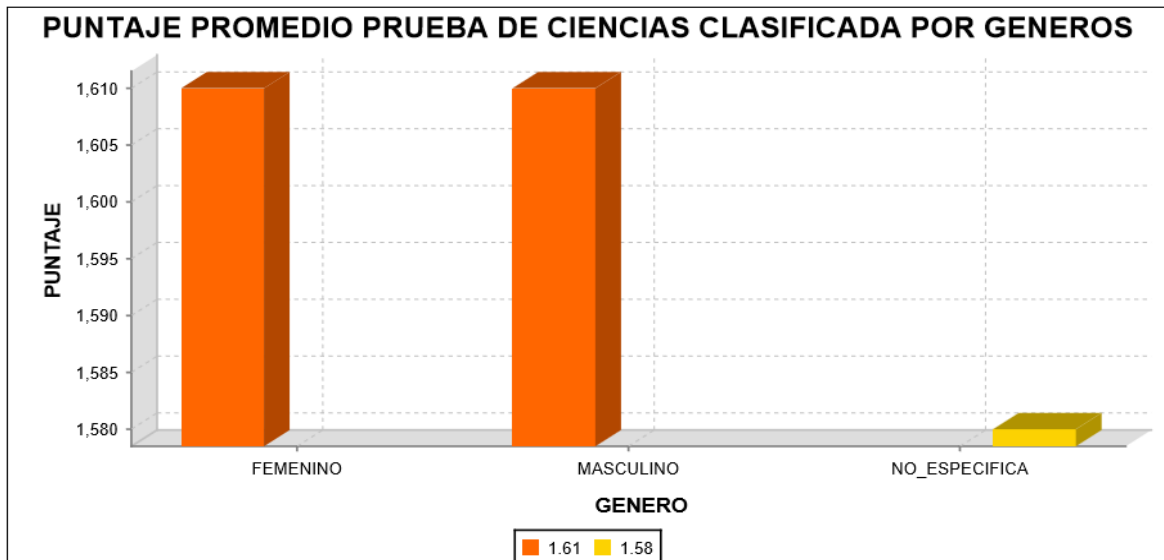
Figura 31. Puntaje promedio de la prueba de Lenguaje clasificados por géneros.



Fuente: Esta investigación.

13. ¿Puntaje promedio de la prueba de Ciencias clasificados por géneros?

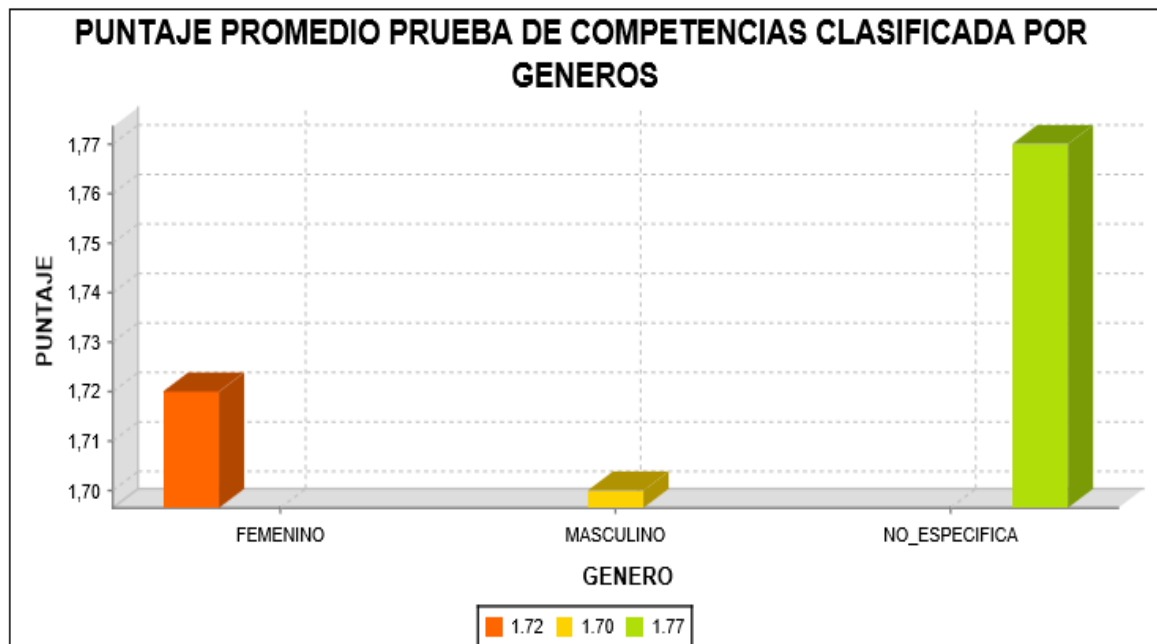
Figura 32. Puntaje promedio de la prueba de ciencias, clasificados por géneros.



Fuente: Esta investigación.

14. ¿Puntaje promedio de la prueba de Competencias clasificados por géneros?

Figura 33. Puntaje promedio de la prueba de Competencias clasificados por géneros.



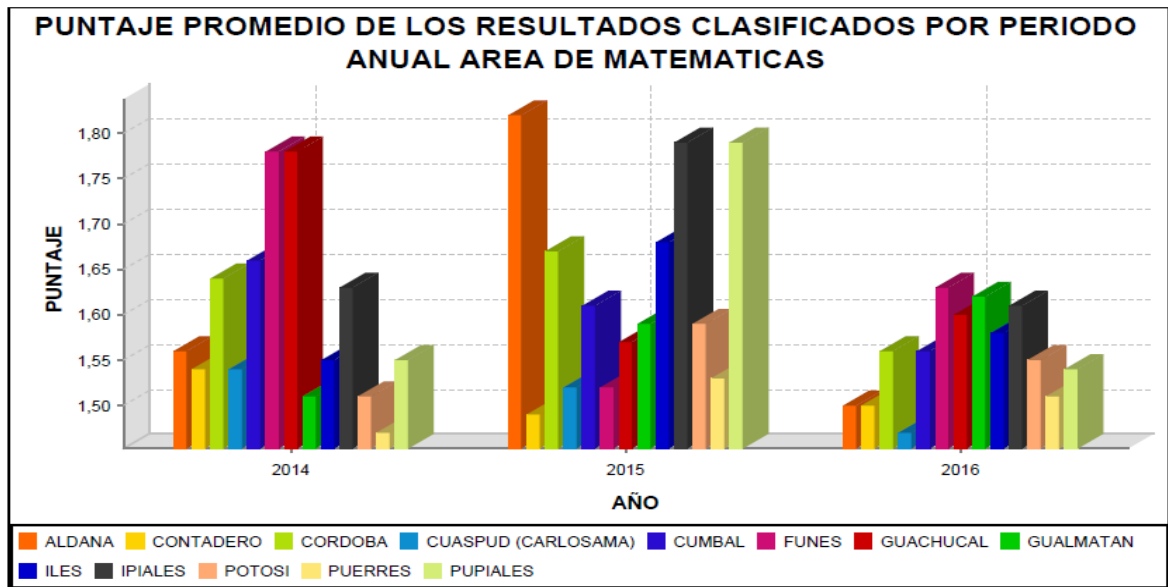
Fuente: Esta investigación.

En las figuras 30, 31, 32 y 33 podemos observar el comportamiento de los estudiantes que presentaron las pruebas saber 5° en los años 2014 al 2016 en las áreas de matemáticas, lenguaje, competencias y ciencias clasificados por género, obteniendo los siguientes resultados:

En el área de matemáticas y lenguaje el género que obtuvo el mejor promedio fue el 'Masculino', para el área de ciencias podemos observar que el género 'Masculino' y 'Femenino' tienen un comportamiento similar. Y finalmente para el área de competencias el 'No_Especificia' tiene el mejor promedio y el 'Masculino' el menor promedio.

15. ¿Puntaje promedio de los resultados clasificados por año área de matemáticas?

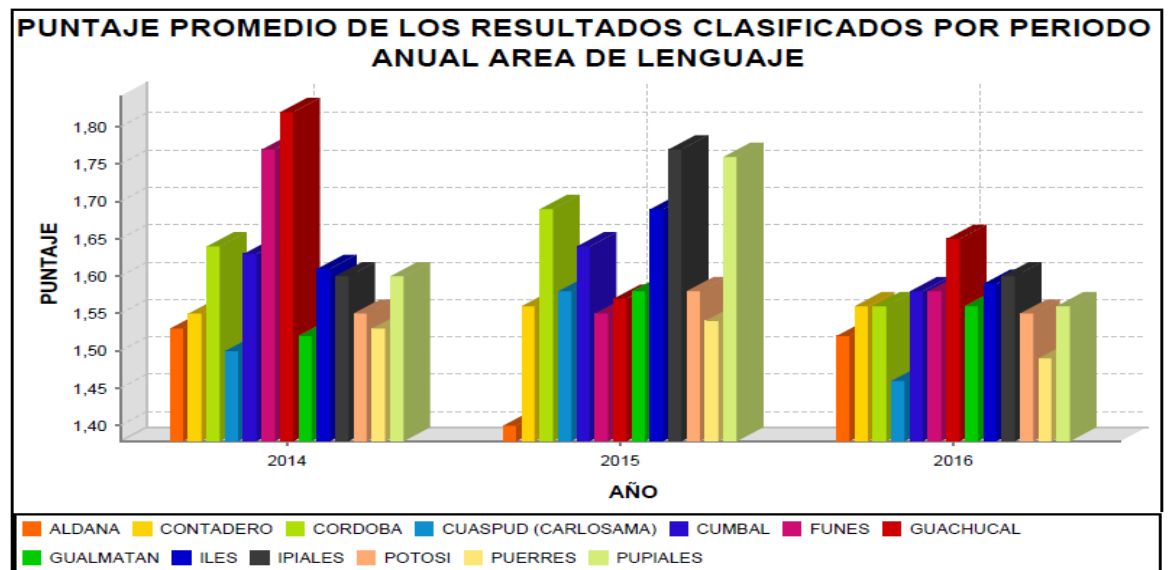
Figura 34. Puntaje promedio de los resultados clasificados por año área de matemáticas.



Fuente: Esta investigación.

16. ¿Puntaje promedio de los resultados clasificados por año área de Lenguaje?

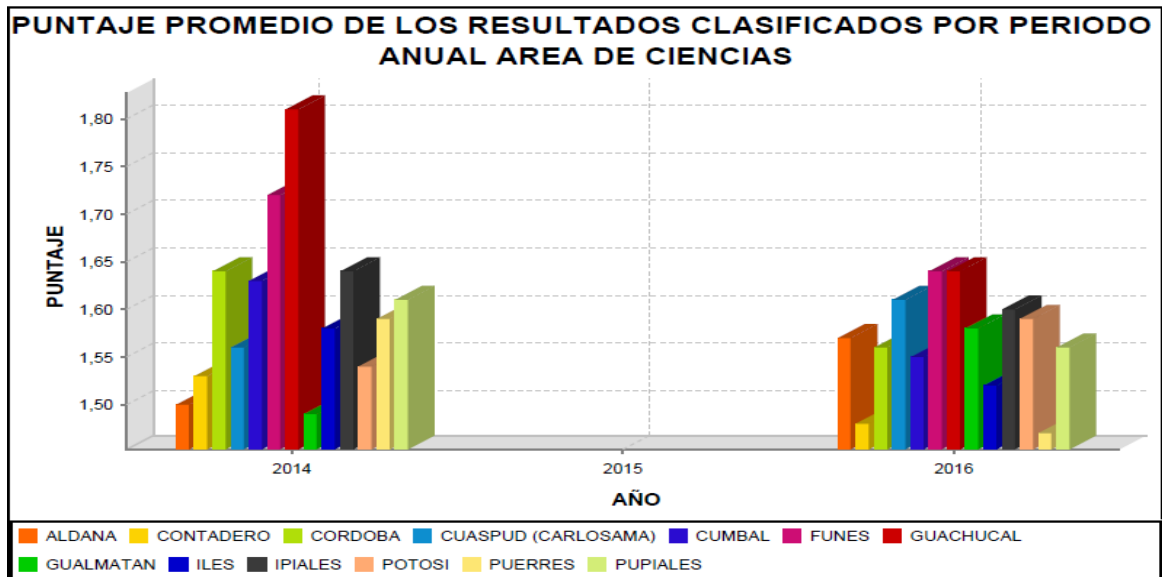
Figura 35. Puntaje promedio de los resultados clasificados por año área de Lenguaje.



Fuente: Esta investigación.

17. ¿Puntaje promedio de los resultados clasificados por año área de Ciencias?

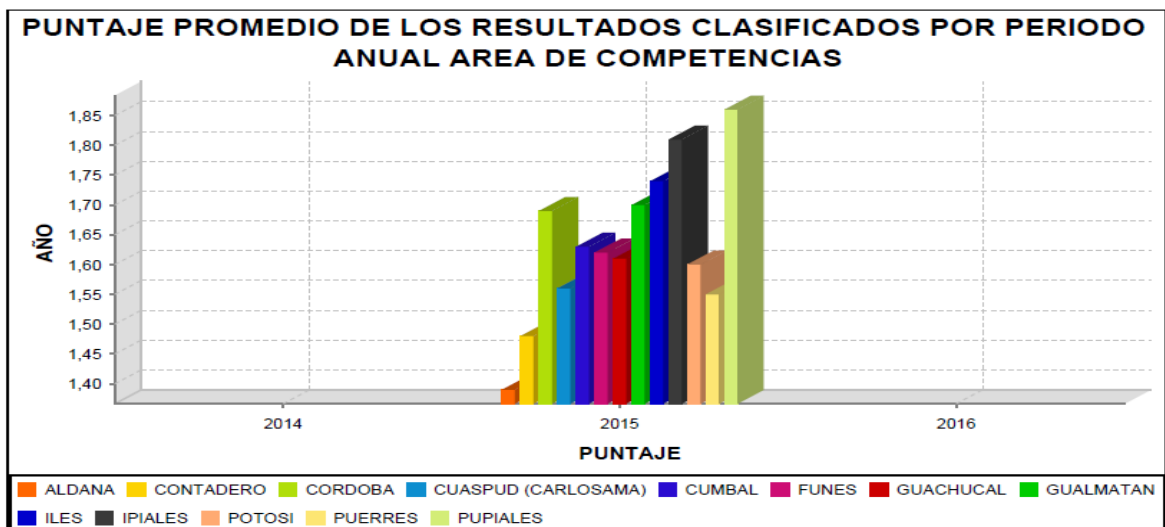
Figura 36. Puntaje promedio de los resultados clasificados por año área de Ciencias.



Fuente: Esta investigación.

18. ¿Puntaje promedio de los resultados clasificados por año área de Competencias?

Figura 37. Puntaje promedio de los resultados clasificados por año área de Competencias.



Fuente: Esta investigación.

En las figuras 34, 35, 36 y 37 podemos observar el comportamiento de los estudiantes que presentaron las pruebas saber 5° en los años 2014 al 2016 en las áreas de matemáticas, lenguaje, competencias y ciencias clasificados por año en los diferentes municipios de la subregión de Obando, obteniendo los siguientes resultados:

Tabla 25. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las 4 áreas, clasificado por periodo anual.

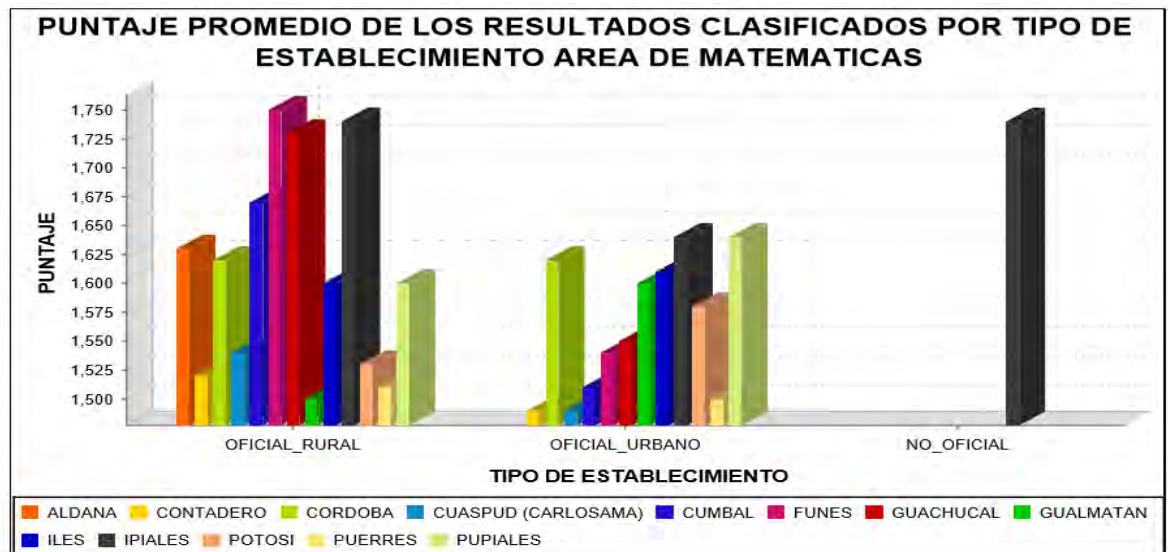
Período/año		2014	2015	2016
Matemáticas	Mayor	Funes Guachucal	Aldana	Funes
	Menor	Puerres	Contadero	Cuaspud(Carlosama)
Lenguaje	Mayor	Guachucal	Guachucal	Ipiales
	Menor	Cuaspud(Carlosama)	Aldana	Cuaspud(Carlosama)
Ciencias	Mayor	Guachucal	-	Funes Guachucal
	Menor	Gualmatan	-	Puerres
Competencias	Mayor	-	Pupiales	-
	Menor	-	Aldana	-

Fuente: Esta investigación.

De la tabla 25 podemos concluir que uno de los municipios con mejores promedios es Guachucal, Además Funes tiene promedios aceptables en el área de Matemáticas, por su parte Puerres, Cuaspud (Carlosama) y Aldana tiene tendencia a los promedios más bajos en algunas de las Áreas.

19. ¿Puntaje promedio de los resultados de Matemáticas clasificados por tipo de establecimiento?

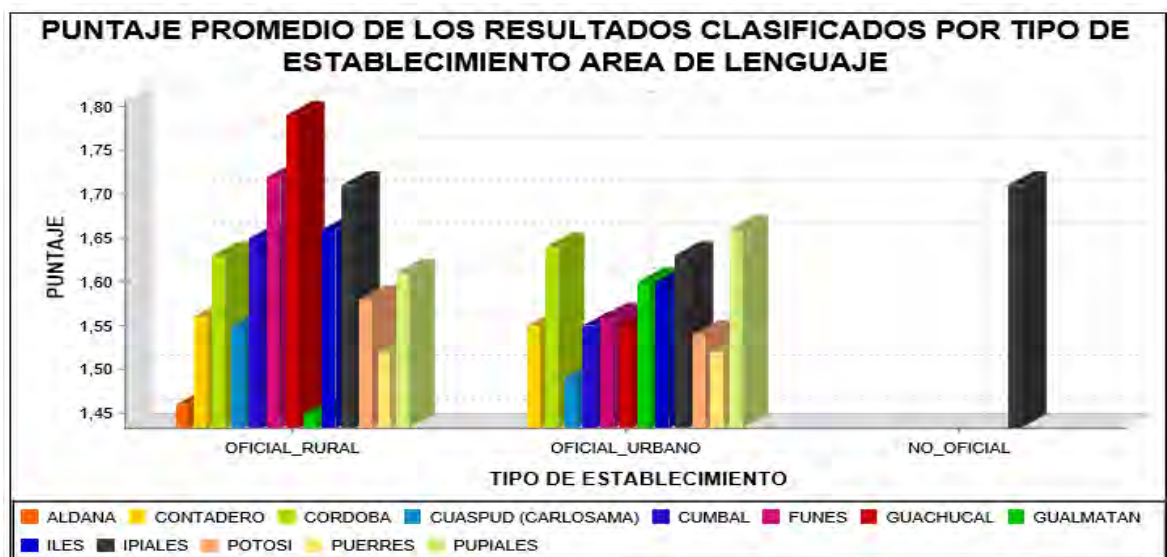
Figura 38. Puntaje promedio de los resultados de Matemáticas clasificados por tipo de establecimiento.



Fuente: Esta investigación.

20. ¿Puntaje promedio de los resultados de Lenguaje clasificados por tipo de establecimiento?

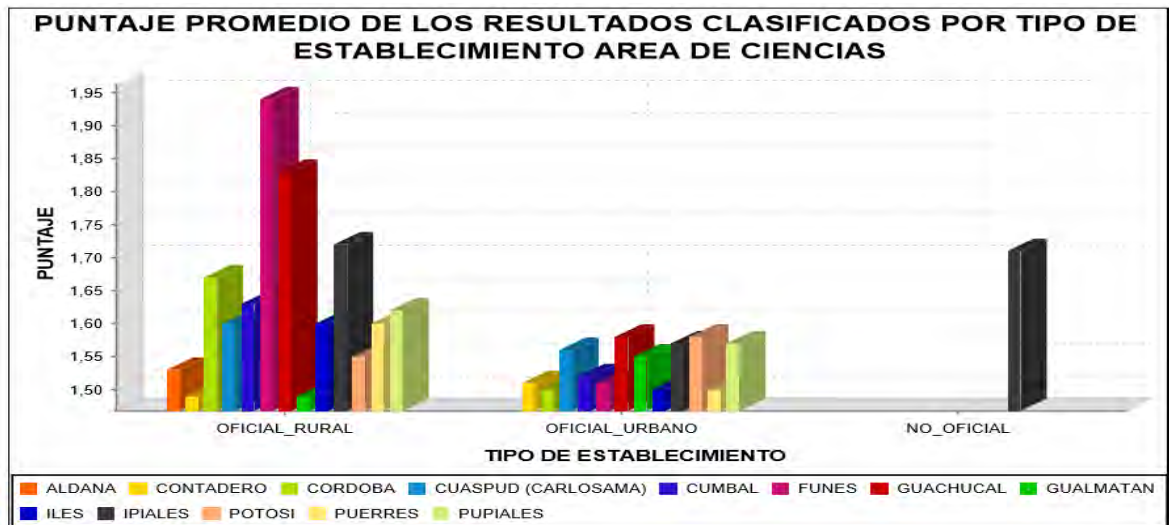
Figura 39. Puntaje promedio de los resultados de Lenguaje clasificados por tipo de establecimiento.



Fuente: Esta investigación.

21. ¿Puntaje promedio de los resultados de Ciencias clasificados por tipo de establecimiento?

Figura 40. Puntaje promedio de los resultados de Ciencias clasificados por tipo de establecimiento.



Fuente: Esta investigación.

22. ¿Puntaje promedio de los resultados de Competencias clasificados por tipo de establecimiento?

Figura 41. Puntaje promedio de los resultados de Competencias clasificados por tipo de establecimiento.



Fuente: Esta investigación.

En las figuras 38, 39, 40 y 41 podemos observar el comportamiento de los estudiantes que presentaron las pruebas saber 5° en los años 2014 al 2016 en las áreas de matemáticas, lenguaje, competencias y ciencias clasificados por tipo de establecimiento en los diferentes municipios de la subregión de Obando, obteniendo los siguientes resultados:

Tabla 26. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las 4 áreas, clasificado por tipo de establecimiento.

Áreas	Tipo IE		Oficial_Rural	Oficial_Urbano	No_Oficial
	Mayor	Menor			
Matemáticas	Mayor		Funes	Ipiales Pupiales	-
	Menor		Gualmatan	Aldana Cuaspud(Carlosama)	-
Lenguaje	Mayor		Guachucal	Pupiales	-
	Menor		Gualmatan	Cuaspud(Carlosama)	-
Ciencias	Mayor		Funes	-	-
	Menor		Gualmatan Contadero	Todos	-
Competencias	Mayor		Ipiales	Pupiales	-
	Menor		Aldana	Contadero	-

Fuente: Esta investigación.

Con base en la tabla 26 podemos concluir que Gualmatan, Aldana y Cuaspud(Carlosama) tienen promedios significativamente bajos con respecto a los diferentes municipios de la subregión de Obando, lo cual pone en alerta a las instituciones Educativas y los entes encargados del mejoramiento de la calidad académica. Por su parte cabe resaltar que el único municipio que cuenta con establecimientos educativos No_Oficiales es Ipiales, que además tiene un promedio regular, el cual con estrategias de mejoramiento se puede aumentar.

23. ¿Puntaje promedio de los resultados de Matemáticas clasificados por zona?

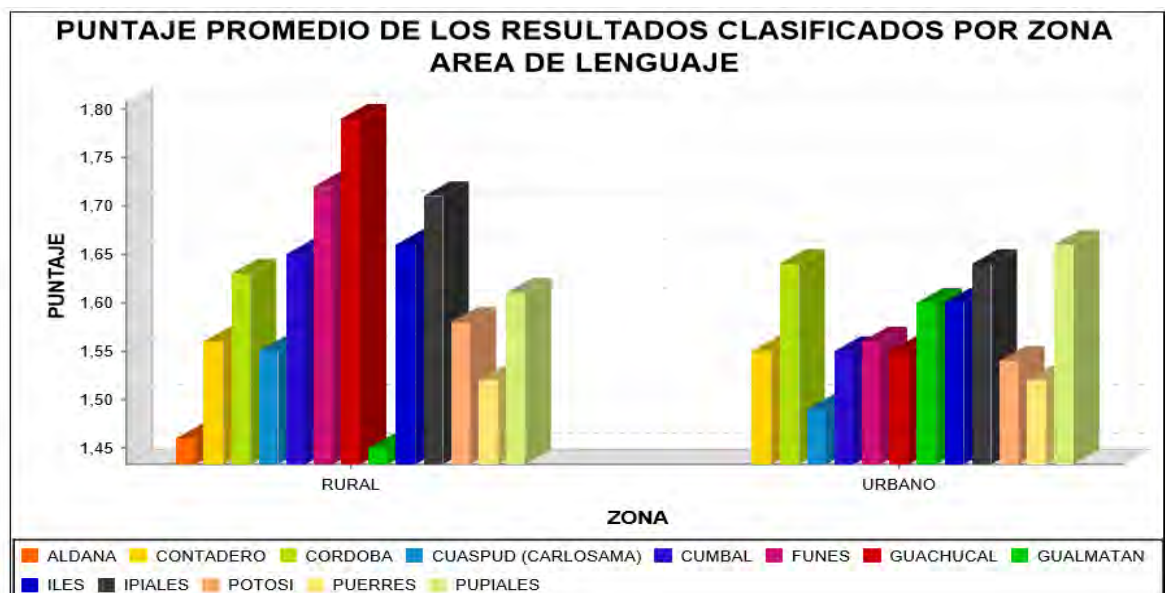
Figura 42. Puntaje promedio de los resultados de Matemáticas clasificados por zona.



Fuente: Esta investigación.

24. ¿Puntaje promedio de los resultados de Lenguaje clasificados por zona?

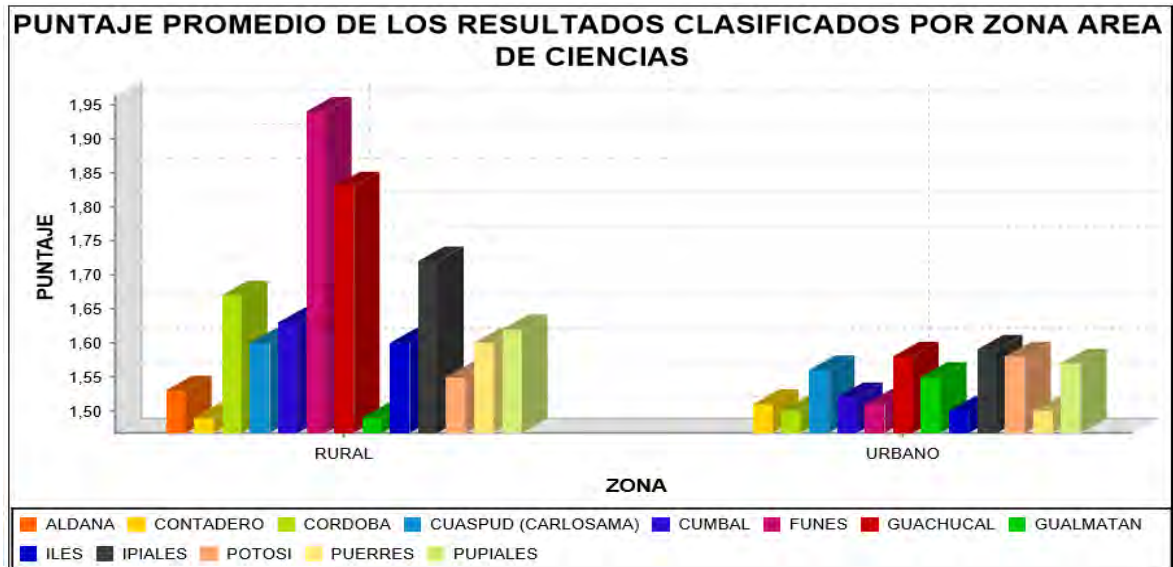
Figura 43. Puntaje promedio de los resultados de Lenguaje clasificados por zona.



Fuente: Esta investigación.

25. ¿Puntaje promedio de los resultados de Ciencias clasificados por zona?

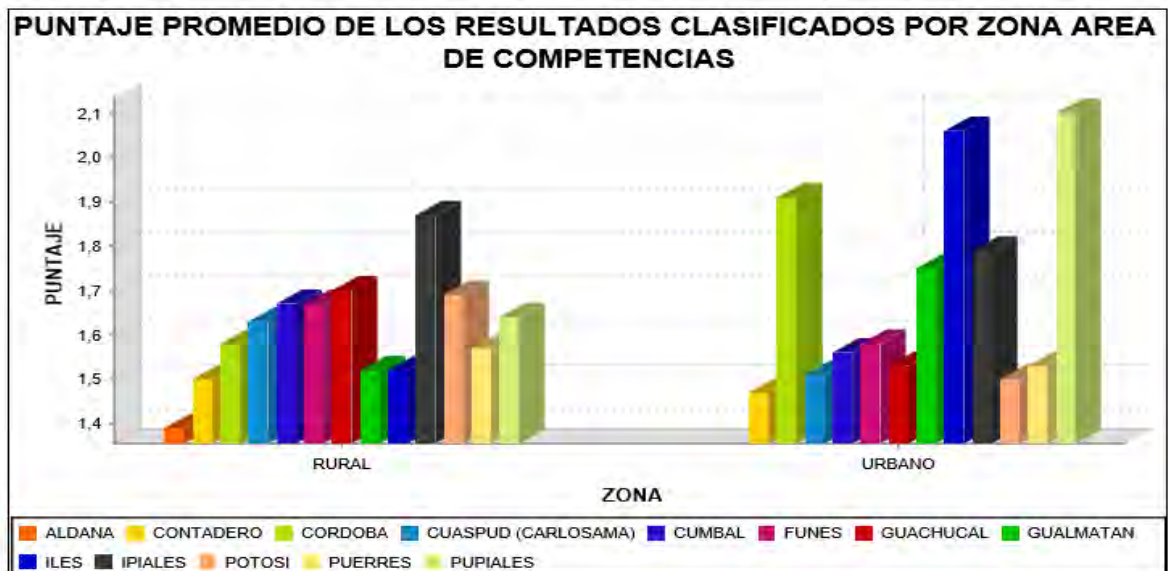
Figura 44. Puntaje promedio de los resultados de Ciencias clasificados por zona.



Fuente: Esta investigación.

26. ¿Puntaje promedio de los resultados de Competencias clasificados por zona?

Figura 45. Puntaje promedio de los resultados de Competencias clasificados por zona.



Fuente: Esta investigación.

En las figuras 42, 43, 44 y 45 podemos observar el comportamiento de los estudiantes que presentaron las pruebas saber 5° en los años 2014 al 2016 en las áreas de matemáticas, lenguaje, competencias y ciencias clasificados por zona en los diferentes municipios de la subregión de Obando, obteniendo los siguientes resultados:

Para las áreas de matemáticas, lenguaje y ciencias podemos concluir que la zona rural tiene mejores promedios en comparación con la zona urbana, caso contrario sucede para el área de competencias, por consiguiente, se especifica los resultados en la tabla 27:

Tabla 27. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las 4 áreas, clasificado por zona.

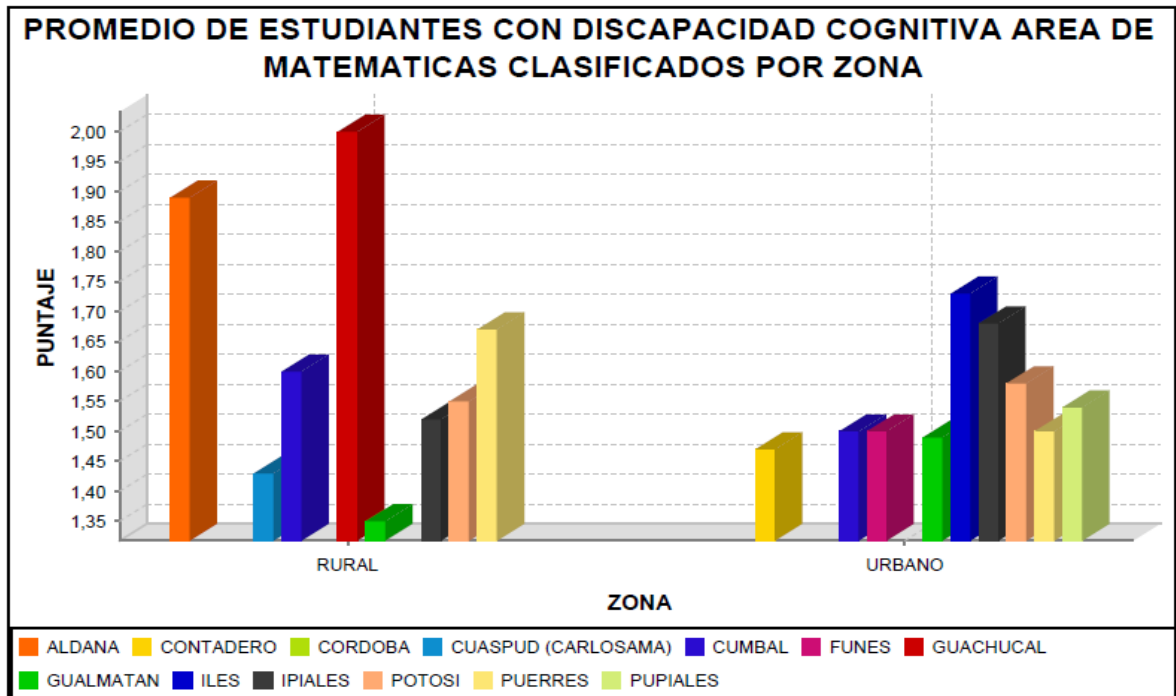
Áreas	Zona		Rural	Urbano
	Mayor	Menor		
Matemáticas	Mayor		Funes	Ipiales
	Menor		Gualmatan	Contadero Cuaspud(Carlosama)
Lenguaje	Mayor		Guachucal	Pupiales
	Menor		Gualmatan	Cuaspud(Carlosama)
Ciencias	Mayor		Funes	-
	Menor		Gualmatan Contadero	Todos
Competencias	Mayor		Ipiales	Pupiales
	Menor		Aldana	Contadero

Fuente: Esta investigación.

En la tabla 27 podemos observar que Gualmatan y Cuaspud (Carlosama) tienen promedios significativamente bajos de acuerdo a los resultados generales clasificados por zona geográfica. Por otra parte, cabe resaltar que todos los municipios de la zona urbana tienen un comportamiento muy bajo en el área de ciencias, lo cual sería muy importante revisar con los diferentes entes encargados de mejorar la calidad educativa.

27. ¿Puntaje de estudiantes con discapacidad cognitiva área de matemáticas clasificados por zona?

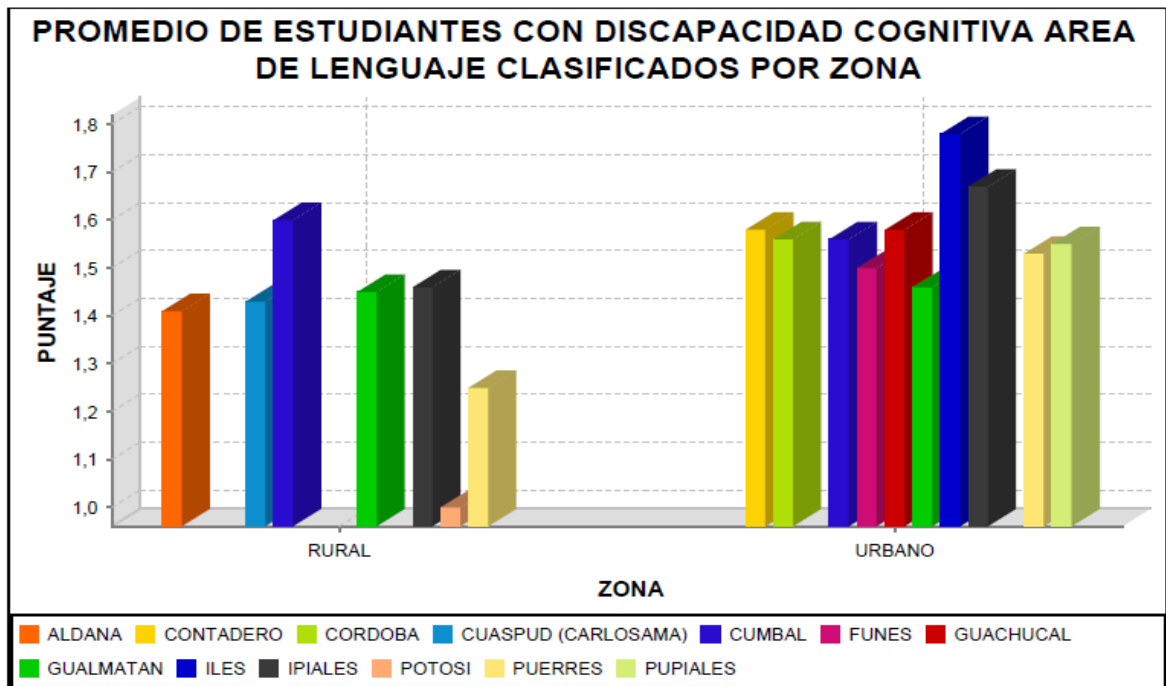
Figura 46. Puntaje de estudiantes con discapacidad cognitiva área de matemáticas clasificados por zona.



Fuente: Esta investigación.

28. ¿Puntaje de estudiantes con discapacidad cognitiva área de Lenguaje clasificados por zona?

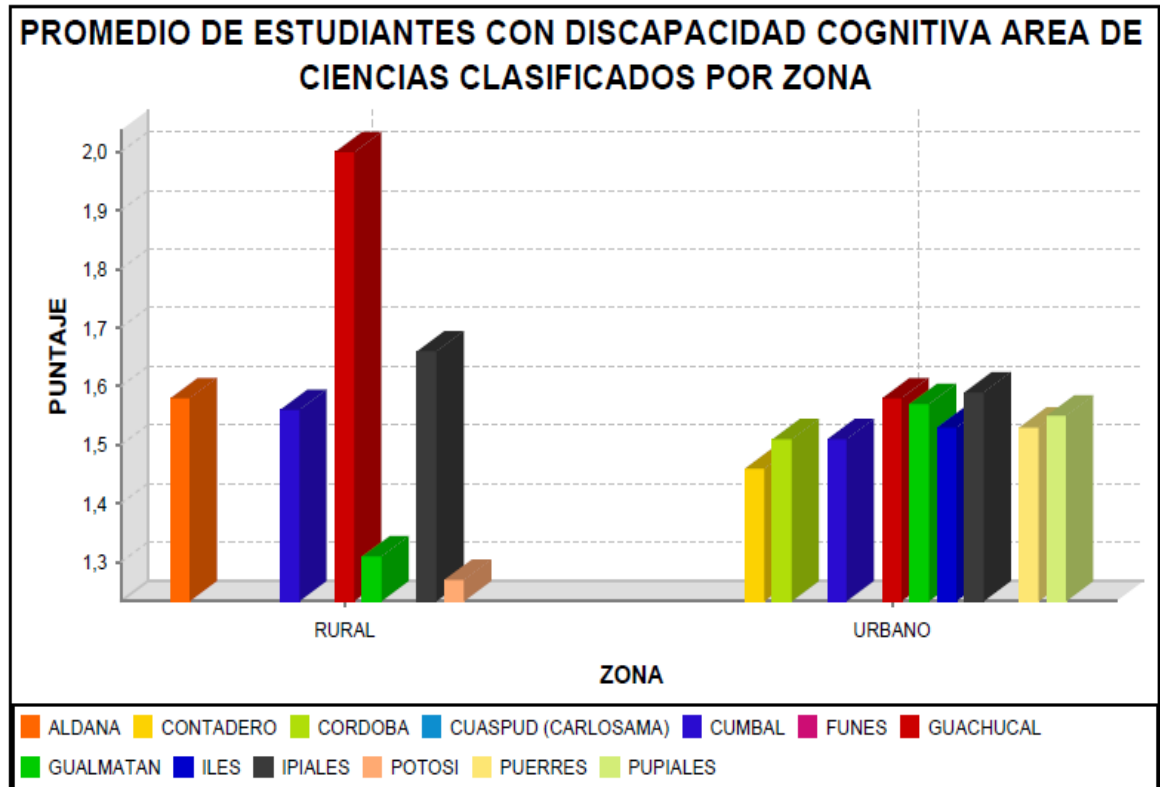
Figura 47. Puntaje de estudiantes con discapacidad cognitiva área de lenguaje clasificados por zona.



Fuente: Esta investigación.

29. ¿Puntaje de estudiantes con discapacidad cognitiva área de Ciencias clasificados por zona?

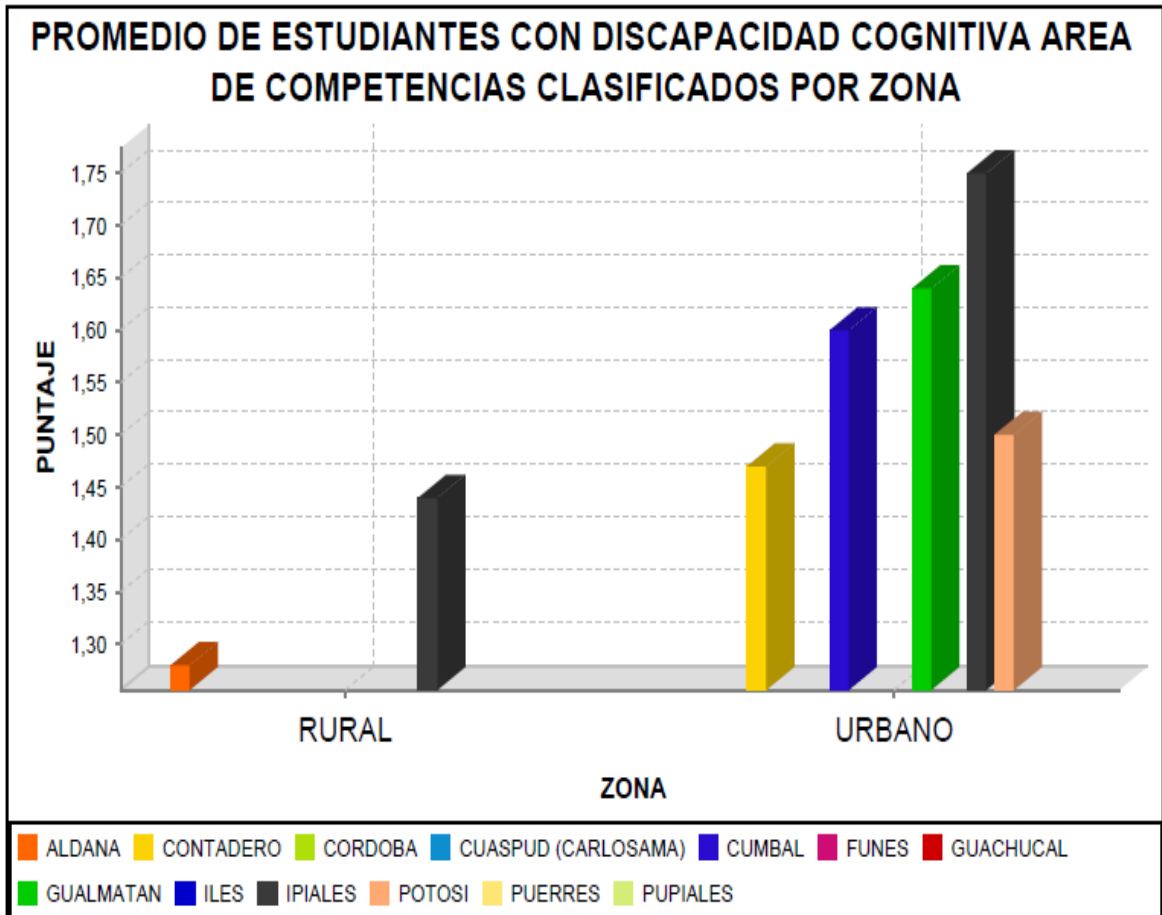
Figura 48. Puntaje de estudiantes con discapacidad cognitiva área de ciencias clasificados por zona.



Fuente: Esta investigación.

30. ¿Puntaje de estudiantes con discapacidad cognitiva área de competencias clasificados por zona?

Figura 49. Puntaje de estudiantes con discapacidad cognitiva área de competencias clasificados por zona.



Fuente: Esta investigación.

En las figuras 46, 47, 48 y 49 podemos observar el comportamiento de los estudiantes con discapacidad cognitiva que presentaron las pruebas saber 5° en los años 2014 al 2016 en las áreas de matemáticas, lenguaje, competencias y ciencias clasificados por zona, en los diferentes municipios de la subregión de Obando, obteniendo los siguientes resultados:

Tabla 28. Descripción municipios con mayor y menor promedio en las pruebas saber 5° en las 4 áreas, clasificado por discapacidad cognitiva en las zonas geográficas.

Áreas	Zona	Rural	Urbano
	Matemáticas	Mayor	Guachucal
Menor		Gualmatan	-
Lenguaje	Mayor	Cumbal	Cumbal
	Menor	Potosí	-
Ciencias	Mayor	Guachucal	-
	Menor	Potosí	Todos
Competencias	Mayor	Ipiales	Ipiales
	Menor	Aldana	Contadero

Fuente: Esta investigación.

De la tabla 28 podemos concluir que Ipiales, Cumbal y Guachucal tienen promedios mayores con respecto al resto de los municipios de la subregión de Obando. Por otra parte, Potosí, Aldana y el Contadero tienen promedios bajos. Cabe resaltar que la mayoría de municipios obtuvieron promedios realmente bajos con respecto a los resultados a nivel nacional.

CONCLUSIONES Y RECOMENDACIONES

El objetivo de crear un Datamart para el almacenamiento histórico de los resultados de las pruebas saber 5, como una herramienta para soportar las decisiones de los entes gubernamentales y directivos educativos de la subregión de Obando se cumplió en su totalidad.

Como resultado de ello la subregión de Obando cuenta con un Datamart de las pruebas saber 5 de los estudiantes de esta subregión que le permitirá conjuntamente con la herramienta Pentaho realizar un análisis por múltiples dimensiones o variables y visualizar sus resultados de manera gráfica, facilitando la toma de decisiones.

En este proyecto, la etapa que más consumió recursos y tiempo fue la de extracción, transformación y carga de los datos de las bases de datos del ICFES al Datamart, por la mala calidad de los datos que implicó su limpieza y transformación. Lo anterior se justifica si se tiene en cuenta que son las fases fundamentales que pueden garantizar el éxito del proyecto.

Para obtener resultados confiables es fundamental realizar un análisis serio de la calidad inicial de los datos, con el fin de establecer los procedimientos necesarios para realizar la limpieza y estandarización de los mismos.

Es necesario documentar cada una de las fases del proyecto y generar los metadatos correspondientes a fin de poder replicar el proceso en trabajos futuros de la misma naturaleza.

En consideración a que una de las políticas establecidas por el gobierno nacional para medir la calidad de las IE son los resultados de las pruebas Saber 5°, se recomienda que las Instituciones Educativas dediquen mayor tiempo, esfuerzo y recursos para establecer políticas serias de mejoramiento.

Como trabajo futuro se propone la construcción de un aplicativo web que se conecte al Datamart construido y que permita a los municipios de la subregión de Obando actualizar la información de los resultados de las pruebas saber 5 en los años por venir.

REFERENCIAS

- [1] H. Fernández. Como Interpretar la Evaluación Pruebas Saber. Bogotá, Colombia: Ministerio de Educación Nacional, 2005.
- [2] ICFES. Sistema Nacional de Evaluación Estandarizada de la Educación Alineación del examen SABER 11°. Lineamientos generales 2014 – 2. ISBN: 978-958-11-0630-1. Bogotá, Colombia, 2014.
- [3] ICFES. Pruebas Saber 3º, 5º y 9º. Lineamientos para las aplicaciones muestral y censal 2015. Bogotá, Colombia, 2015.
- [4] ICFES (2016). Consulta de resultados, pruebas saber 3º 5º 9º [Online]. Available: <http://www2.icfesinteractivo.gov.co/ReportesSaber359/consultaReporteEntidadTerritorial.jspx>.
- [5] T. Pérez. (2006). Plan Nacional Decenal de Educación 2006 – 2016. Compendio general. (Página 16). [Online]. Available: www.plandecenal.edu.co/html/1726/articles-198148_archivo_pdf.pdf.
- [6] G. R. Rivadera. La metodología de Kimball para el diseño de almacenes de datos (Datawarehouses). Buenos Aires, Argentina, 2010.
- [7] PostgreSQL. (2016, 14 Oct). Todo Sobre PostgreSQL [Online]. Available: http://www.postgresql.org.es/sobre_postgresql.
- [8] Dataprix.com. (2016). ¿Qué es un Data Warehouse? | Dataprix TI. [online] Available at: <http://www.dataprix.com/que-es-un-datawarehouse> [Accessed 14 Oct. 2016].
- [9] Sinnexus.com. (2016). Datawarehouse. [online] Available at: http://www.sinnexus.com/business_intelligence/datawarehouse.aspx [Accessed 14 Oct. 2016].
- [10] Sinnexus.com. (2016). Datamart. [online] Available at: http://www.sinnexus.com/business_intelligence/datamart.aspx [Accessed 21 May. 2017].

- [11] R. Fuyol's (2017, 21 May) Blog. ¿Qué es Pentaho y para qué sirve?. [online]. Available: <https://rodryfuyol.wordpress.com/que-es-pentaho-y-para-que-sirve/>. 2017.
- [12] G. R. Rivadera, La metodología de Kimball para el diseño de almacenes de datos (Datawarehouses), en Cuadernos de la Facultad de Ingeniería e Informática Número 5, Universidad Católica de Salta, 2010.
- [13] E. Malinowski y E. Zimányi, Advanced data warehouse design. From Conventional to Spatial and Temporal Applications. Berlin: Springer-Verlag, 2008.
- [14] B. Inmon, Building the Data Warehouse, John Wiley & Sons, 2002.
- [15] M. Reyes y P. Rosales. Trabajo de grado para optar al título de Ingeniero en Ciencias y Sistemas. “Desarrollo de un datamart de información académica de estudiantes de la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería de la USAC”. Guatemala, 2007
- [16] A. Rojas. Tesis para optar al título profesional de Ingeniero de Computación y Sistemas, “Implementación de un datamart como solución de inteligencia de negocios, bajo la metodología de Ralph Kimball para optimizar la toma de decisiones en el Departamento de Finanzas de la Contraloría General de la República”. Chiclayo, 2014.
- [17] R. Kimball y A. Ross, J. Wiley & Sons The data warehouse toolkit, Third. Indianapolis: 2013.
- [18] Sinnexus.com. (2016). Modelo dimensional. [online] Available at: www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.0/com.ibm.datatools_dimensional.ui.doc/topics/c_dm_star_schemas.html
- [19] Unesco. Factores asociados al logro cognitivo de los estudiantes en América Latina y el Caribe. Santiago: Oficina Regional de Educación de la UNESCO para América Latina y el Caribe, 2015.
- [20] J. S. Coleman, E. Campbell, C. Hobson, J. McFarland, A. Mood, F. Weinfeld, and R. York. Equality of Educational Opportunity. National Center for Educational Statistics. Report number OE38001. US Government Printing Office. Washington, USA, 1966.

- [21] C. Goldin, L. Katz, & I. Kuziemko. The homecoming of american college women: the reversal of the college gender gap, 2006.
- [22] M. Niederle, & L. Vesterlund. Explaining the Gender Gap in Math Test Scores: he Role of Competition. *The Journal of Economic Perspectives*, 24(2), 129-144. [Online]. Available: <http://web.stanford.edu/~niederle/NV.JEP.pdf>, 2010
- [23] Unesco. Recomendaciones de políticas educativas para América Latina en base al TERCE. Santiago: Unesco, 2016.
- [24] Unesco. ¿Es la repitencia efectiva? (Vol. 1). Santiago: Oficina Regional de Educación para América Latina y el Caribe, 2015.
- [25] M. T. Celis, Ó. A. Jiménez, & J. F. Jaramillo. ¿Cuál es la brecha de la calidad educativa en Colombia en la educación media y en la superior? En *Estudios sobre calidad de la educación en Colombia* (pp. 67-98). Bogotá: Instituto Colombiano para la Evaluación de la Educación ICFES, 2012.
- [26] E. Chaux. (2009). Citizenship Competencies in the Midst of a Violent Political Conflict: The Colombian Educational Response. *Harvard Educational Review*, 79(1), 84–93. [Online]. Available: <https://doi.org/10.17763/haer.79.1.d2566q027573h219>, 2009.
- [27] E. M. Pomerantz, E. A. Moorman, & S. D. Litwack. The How, Whom, and Why of Parents' Involvement in Children's Academic Lives: More Is Not Always Better. *Review of Educational Research*, 77(3), 373-410. doi:10.3102/003465430305567, 2007.
- [28] Unesco. Guidelines for inclusion: ensuring access to education for all. Paris: United Nations Educational, Scientific and Cultural Organization, 2005.
- [29] K. Dostes y Scheneider. self-concept attributional beliefs and school achievement: a longitudinal analysis. *Contemporary educational psychology* 19(2) 199-216, 1994.
- [30] R. M. Ryan & E. L. Deci. *Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions*; University of Rochester, 2000.

- [31] OCDE. Learners for life. Student approaches to learning. Results from PISA 200. Paris: OECD Publishing, 2003.
- [32] OCDE. Pisa 2009 Results: Learning to Learn - Student Engagement, Strategies and Practices, 2010b.
- [33] OECD. PISA 2015 Assessment and Analytical Framework: OECD Publishing, 2016a.
- [34] J. Ospina. La motivación, motor del aprendizaje, Rev. Cienc. Salud. Bogotá (Colombia) 4 (Especial): 158-160, 2006.
- [35] Weinstein, C. E., Husman, J., & Dierking, D. R. (2000). Self-regulation interventions with a focus on learning strategies. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), Handbook of selfregulation (pp. 727–747). San Diego, CA: Academic Press.
- [36] J. Chubb. The profit motive. The private can the public. Education Next, 1 2001.
- [37] S. Ball. Privatizaçoa da educaçao e novas subjetividades: contornos e esdobramientos das políticas (pós) neoliberais. Revista Brasileira de Educação, 18(53), 457-466, 2012.
- [38] M. Crenshaw. The relationships among school size, school climate variables, and achievement ratings in South Carolina high schools: A conceptual model. University of South Carolina, Columbia.2003.
- [39] D. Lamdin. Sting for effect of school size on student achievement within a school district. Education Economics, 3, 33-42, 1995.
- [40] A. Ajani, & R. Akinyele. Effects of Student-Teacher Ratio on Academic Achievement of Selected Secondary School Students in Port Harcourt Metropolis, Nigeria. Journal of Education and Practice, 5(24), 100-106, 2014.
- [41] K. Stevenson. Elementary school capacity: What size is the right size? CEFPP Journal, 33(4), 10-14, 1996.

- [42] E. Duflo, P. Dupas. & M. Kremer. Peer effects, Teacher Incentives and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*. 101(5), pp. 1739-1774, 2011.
- [43] N. K. Bowen, & G. L. Bowen. Effects of Crime and Violence in Neighborhoods and Schools on the School Behavior and Performance of Adolescents. *Journal of Adolescent Research*, 14(3), 319-342. Doi: 10.1177/0743558499143003, 1999.
- [44] OECD. *Starting Strong III: a quality toolbox for early childhood education and care*. Paris: OECD Publishing, 2012.
- [46] C. Holmes. Grade Level Retention Effects: A Meta-Analysis of Research Studies in L.A Shepard and M. L. Smith. (Eds), *Flunking Grades: Research and Policies on Retention*. The Falmer Press, 1989.
- [47] M. Roderick. Grade Retention and School Dropout: Investigating the Association. *American Educational Research Journal*. Vol. 31(4). Pp. 729-759, 1994.
- [48] OECD (2015). *Students, Computers and Learning: Making the Connection*. PISA. OECD [Online]. Available: <http://dx.doi.org/10.1787/9789264239555-en>.
- [49] F. Reimers, DeShano da C. Silva, & E. Treviño (2006). *Where is the Education in Conditional Cash Transfer in Education?* UNESCO Institute for Statistics. [Online]. Available: <http://www.uis.unesco.org/Library/Documents/WP1-06-en.pdf>.
- [50] J. Murillo. *Investigación iberoamericana sobre eficacia escolar*. Bogotá: Convenio Andrés Bello, 2007.
- [51] K. Leithwood. Leadership for school restructuring. *Educational administration quarterly*, 30(4), 498-518, 1994.
- [52] R. Elmore. *Mejorando la escuela desde la sala de clases*. Área de Educación Fundación Chile, 2010.
- [53] J. Cohen, L. McCabe, N. M. Michelli & T. Pickeral. School climate: Research, policy, practice, and teacher education. *Teachers College Record*, 111(1), 180-213, 2009.

- [54] K. Lester, J. Garofalo & D. Kroll. Self-Confidence, Interest, Beliefs, and Metacognition: Key Influences on Problem-Solving Behavior. In D.B. McLeod & V. M. Adams (Eds.), *Affect and mathematical problem solving* (75-88). Springer New York, 1989.
- [55] G. Ferrer. *Sistemas de evaluación de aprendizajes en América Latina: balance y desafíos*: PREAL, 2006.
- [56] J. Westheimer, & J. Kahne. What kind of citizen? The politics of educating for democracy. *American Educational Research Journal*, 41(2), 237-269, 2004.
- [57] V. Espínola. *Educación para la ciudadanía y la democracia para un mundo globalizado: una perspectiva comparativa*. Washington: Banco Interamericano de Desarrollo, 2005.
- [58] ICFES. *Alineación del examen SABER 11°. Sistema Nacional de Evaluación Estandarizada de la Educación*, Instituto Colombiano para la Evaluación de la Educación (ICFES). Bogotá, Colombia, 2013.
- [59] D. S. Rychen & L. H. Salganik. Traducción al español: *Las competencias clave para el bienestar personal, económico y social* (1ª ed. en español), 2006.
- [60] L. Gomes & Y. Jaramillo. *Descubrimiento de factores asociados al desempeño en las pruebas saber 5 con técnicas descriptivas de minería de datos*, San Juan de Pasto, Colombia, 2017.
- [61] H. Rodríguez. El paradigma de las competencias hacia la educación superior. *Revista Facultad de Ciencias Económicas*. Universidad Militar Nueva Granada. V. XV, N. 1, 145165, 2007.
- [62] M. Torrado. *Educación para el desarrollo de las competencias: Una propuesta para reflexionar*. En *Competencias y proyecto pedagógico*. Universidad Nacional, Bogotá, Colombia, 2000.
- [63] ICFES. *Marco de factores Asociados Saber 3°, 5° y 9° 2016*. Instituto Colombiano para la Evaluación de la Educación (ICFES). ISBN de la versión digital: En trámite. Bogotá, Colombia, 2011.
- [64] Unesco. *Factores Asociados. Informe de resultados Tercer Estudio Regional Comparativo y Explicativo*. Santiago: Oficina Regional de Educación para América

Latina y el Caribe (OREALC/Unesco Santiago). [Online]. Available: <http://unesdoc.unesco.org/images/0024/002435/243533s.pdf>, 2015b

[65] ICFES. Informe técnico Saber 5° y 9° 2009. Instituto Colombiano para la Evaluación de la Educación (ICFES). ISBN de la versión electrónica: 978-958-11-0578-6. Bogotá, Colombia, 2011.

ANEXOS

ANEXO 1. SCRIPTS UTILIZADOS EN LA IMPLEMENTACIÓN

A continuación, se presentan los scripts usados para las distintas fases de esta investigación.

Scripts utilizados para Limpieza de datos:

```
create table entidades_2014_2016_limpio as select * from entidades_2014_2016;
```

```
create table establecimientos_2014_2016_limpio as select * from establecimientos_2014_2016;
```

```
create table municipios_2014_2016_limpio as select * from municipios_2014_2016;
```

```
create table valores_plausibles_2014_2016_limpio as select * from valores_plausibles_2014_2016;
```

Scripts utilizados para la Transformación de Datos:

```
create table entidades_2014_2016_final as select * from entidades_2014_2016_limpio;
```

```
update entidades_2014_2016_final set munexclu='ETC' where munexclu='1';
```

```
update entidades_2014_2016_final set munexclu='DPTO' where munexclu='2';
```

```
update entidades_2014_2016_final set munexclu='SIN_DATO' where munexclu='0';
```

```
create table establecimientos_2014_2016_final as select * from establecimientos_2014_2016_limpio;
```

```
update establecimientos_2014_2016_final set zona='URBANO' where zona='1';
```

```
update establecimientos_2014_2016_final set zona='RURAL' where zona='2';
```

```
update establecimientos_2014_2016_final set sector='OFICIAL' where sector='1';
```

```
update establecimientos_2014_2016_final set sector='NO_OFICIAL' where sector='2';
```

```
update establecimientos_2014_2016_final set tipo_estab='OFICIAL_URBANO'
where tipo_estab='1';
```

```
update establecimientos_2014_2016_final set tipo_estab='OFICIAL_RURAL'
where tipo_estab='2';
```

```
update establecimientos_2014_2016_final set tipo_estab='NO_OFICIAL' where
tipo_estab='3';
```

```
update establecimientos_2014_2016_final set
nombre=upper(translate(nombre,'áéíóúÁÉÍÓÚäëïöüÄËÏÖÜÑ','aeiouAEIOUaeiouA
EIOUN')); // este Script se utilizó para estandarizar caracteres especiales.
```

```
create table municipios_2014_2016_final as select * from
municipios_2014_2016_limpio;
```

```
update municipios_2014_2016_final set
nombre=upper(translate(nombre,'áéíóúÁÉÍÓÚäëïöüÄËÏÖÜÑ','aeiouAEIOUaeiouA
EIOUN'));;
```

```
update municipios_2014_2016_final set tipo='SIN_DATO' where tipo='4'
```

Scripts utilizados para la construcción de los cubos de datos:

```
/****NÚMERO DE ESTUDIANTES TIPO DE ESTABLECIMIENTOS *****/
```

```
SELECT
    muni_nombre,
    ins_tipo_estab,
    SUM(numero_estu)
FROM
    dim_municipios
    JOIN fact_saber5
    USING(muni_id_ente)
    JOIN dim_instituciones
    USING(id_institucion)
GROUP BY
    1,2
ORDER BY
    1,2;
```

/**NUMERO DE ESTUDIANTES POR ZONA GEOGRÁFICA Y GENERO*****/**

```
SELECT
    muni_nombre,
    estu_sexo,
    SUM(numero_estu)
FROM
    dim_municipios
    JOIN fact_saber5
        USING(muni_id_ente)
    JOIN dim_estudiante
        USING(id_estudiante)
GROUP BY
    1,2
ORDER BY
    1,2;
```

/**RESULTADO POR MUNICIPIOS PRUEBA DE LENGUAJE ***/**

```
SELECT
    muni_nombre,
    ROUND(AVG(puntaje_lenguaje),2) AS PROMEDIO
FROM
    dim_municipios
    JOIN
    fact_saber5
        USING(muni_id_ente)
GROUP BY
    1
ORDER BY
    1;
```

/**RESULTADO POR MUNICIPIOS PRUEBA DE MATEMÁTICAS***/**

```
SELECT
    muni_nombre,
    ROUND(AVG(puntaje_matematicas),2) AS promedio
FROM
    dim_municipios
```

```
JOIN
  fact_saber5
  USING(muni_id_ente)
GROUP BY
  1
ORDER BY
  1;
```

/RESULTADO POR MUNICIPIOS PRUEBA DE CIENCIAS**/**

```
SELECT
  muni_nombre,
  ROUND(AVG(puntaje_ciencias),2) AS promedio
FROM
  dim_municipios
  JOIN
  fact_saber5
  USING(muni_id_ente)
GROUP BY
  1
ORDER BY
  1;
```

/RESULTADO POR MUNICIPIOS PRUEBA DE COMPETENCIAS**/**

```
SELECT
  muni_nombre,
  ROUND(AVG(puntaje_competencias),2) AS promedio
FROM
  dim_municipios
  JOIN fact_saber5
  USING(muni_id_ente)
GROUP BY
  1
ORDER BY
  1;
```

/RESULTADO GENERAL POR GENERO DE MATEMÁTICAS**/**

```
SELECT
  estusexo,
  ROUND(AVG(puntaje_matematicas),2) AS promedio
```

```
FROM
    dim_estudiante
    JOIN
    fact_saber5
    USING(id_estudiante)
GROUP BY
    1
ORDER BY
    1;
```

/RESULTADO GENERAL POR GENERO DE LENGUAJE**/**

```
SELECT
    estu_sexo,
    ROUND(AVG(puntaje_lenguaje),2) AS promedio
FROM
    dim_estudiante
    JOIN fact_saber5
    USING(id_estudiante)
GROUP BY
    1
ORDER BY
    1;
```

/RESULTADO GENERAL POR GENERO DE CIENCIAS**/**

```
SELECT
    estu_sexo,
    ROUND(AVG(puntaje_ciencias),2) AS promedio
FROM
    dim_estudiante
    JOIN fact_saber5
    USING(id_estudiante)
GROUP BY
    1
ORDER BY
    1;
```

/RESULTADO GENERAL POR GENERO DE COMPETENCIAS**/**

```
SELECT
    estu_sexo,
```



```

        ROUND(AVG(puntaje_competencias),2) AS promedio
FROM
    dim_estudiante
    JOIN fact_saber5
        USING(id_estudiante)
GROUP BY
    1
ORDER BY
    1;

```

/VISUALIZAR PUNTAJE PROMEDIO DE LOS RESULTADOS
CLASIFICADOS POR GENERO DE LENGUAJE**/**

```

SELECT
    muni_nombre,
    estusexo,ROUND(AVG(puntaje_lenguaje),2) AS promedio
FROM
    dim_municipios
    JOIN fact_saber5
        USING(muni_id_ente)
    JOIN dim_estudiante
        USING(id_estudiante)
GROUP BY
    1,2
ORDER BY
    1,2;

```

/VISUALIZAR PUNTAJE PROMEDIO DE LOS RESULTADOS
CLASIFICADOS POR GENERO DE MATEMÁTICAS**/**

```

SELECT
    muni_nombre,
    estusexo,
    ROUND(AVG(puntaje_matematicas),2) AS promedio
FROM
    dim_municipios
    JOIN fact_saber5
        USING(muni_id_ente)
    JOIN dim_estudiante
        USING(id_estudiante)
GROUP BY

```

```
1,2  
ORDER BY  
1,2;
```

```
/**VISUALIZAR PUNTAJE PROMEDIO DE LOS RESULTADOS  
CLASIFICADOS POR GENERO DE CIENCIAS***/
```

```
SELECT  
    muni_nombre,  
    estu_sexo,  
    ROUND(AVG(puntaje_ciencias),2) AS promedio  
FROM  
    dim_municipios  
    JOIN fact_saber5  
    USING(muni_id_ente)  
    JOIN dim_estudiante  
    USING(id_estudiante)  
GROUP BY  
    1,2  
ORDER BY  
    1,2;
```

```
/**VISUALIZAR PUNTAJE PROMEDIO DE LOS RESULTADOS  
CLASIFICADOS POR GENERO DE COMPETENCIAS***/
```

```
SELECT  
    muni_nombre,  
    estu_sexo,  
    ROUND(AVG(puntaje_competencias),2) AS promedio  
FROM  
    dim_municipios  
    JOIN fact_saber5  
    USING(muni_id_ente)  
    JOIN dim_estudiante  
    USING(id_estudiante)  
GROUP BY  
    1,2  
ORDER BY  
    1,2;
```

/RESULTADOS POR INSTITUCIÓN DEL ÁREA LENGUAJE SEGÚN TIPO DEL ESTABLECIMIENTO**/**

```
SELECT
    muni_nombre,
    ins_tipo_estab,
    ROUND(AVG(puntaje_lenguaje),2) AS promedio
FROM
    dim_municipios
    JOIN fact_saber5
    USING(muni_id_ente)
    JOIN dim_instituciones
    USING(id_institucion)
GROUP BY
    1,2
ORDER BY
    1,2;
```

/RESULTADOS POR INSTITUCIÓN DEL ÁREA MATEMÁTICAS SEGÚN TIPO DEL ESTABLECIMIENTO**/**

```
SELECT
    muni_nombre,
    ins_tipo_estab,
    ROUND(AVG(puntaje_matematicas),2) AS promedio
FROM
    dim_municipios
    JOIN fact_saber5
    USING(muni_id_ente)
    JOIN dim_instituciones
    USING(id_institucion)
GROUP BY
    1,2
ORDER BY
    1,2;
```

/RESULTADOS POR INSTITUCIÓN DEL ÁREA CIENCIAS SEGÚN TIPO DEL ESTABLECIMIENTO**/**

```
SELECT
    muni_nombre,
    ins_tipo_estab,
```

```

        ROUND(AVG(puntaje_ciencias),2) AS promedio
FROM
    dim_municipios
    JOIN fact_saber5
    USING(muni_id_ente)
    JOIN dim_instituciones
    USING(id_institucion)
GROUP BY
    1,2
ORDER BY
    1,2;

```

/RESULTADOS POR INSTITUCIÓN DEL ÁREA COMPETENCIAS SEGÚN TIPO DEL ESTABLECIMIENTO**/**

```

SELECT
    muni_nombre,
    ins_tipo_estab,
    ROUND(AVG(puntaje_competencias),2) AS promedio
FROM
    dim_municipios
    JOIN fact_saber5
    USING(muni_id_ente)
    JOIN dim_instituciones
    USING(id_institucion)
GROUP BY
    1,2
ORDER BY
    1,2;

```

/PROMEDIO POR ZONA Y MUNICIPIO DE LA PRUEBA DE LENGUAJE**/**

```

SELECT
    muni_nombre,
    ins_zona,
    ROUND(AVG(puntaje_lenguaje),2) AS promedio
FROM
    dim_instituciones
    JOIN fact_saber5
    USING(id_institucion)

```

```
        JOIN dim_municipios
        USING(muni_id_ente)
GROUP BY
    1,2
ORDER BY
    1,2;
```

/PROMEDIO POR ZONA Y MUNICIPIO DE LA PRUEBA DE MATEMÁTICAS**/**

```
SELECT
    muni_nombre,
    ins_zona,
    ROUND(AVG(puntaje_matematicas),2) AS promedio
FROM
    dim_instituciones
    JOIN fact_saber5
    USING(id_institucion)
    JOIN dim_municipios
    USING(muni_id_ente)
GROUP BY
    1,2
ORDER BY
    1,2;
```

/PROMEDIO POR ZONA Y MUNICIPIO DE LA PRUEBA DE CIENCIAS**/**

```
SELECT
    muni_nombre,
    ins_zona,
    ROUND(AVG(puntaje_ciencias),2) AS promedio
FROM
    dim_instituciones
    JOIN fact_saber5
    USING(id_institucion)
    JOIN dim_municipios
    USING(muni_id_ente)
GROUP BY
    1,2
ORDER BY
```

1,2;

/PROMEDIO POR ZONA Y MUNICIPIO DE LA PRUEBA DE
COMPETENCIAS**/**

```
SELECT
    muni_nombre,
    ins_zona,
    ROUND(AVG(puntaje_competencias),2) AS promedio
FROM
    dim_instituciones
    JOIN fact_saber5
    USING(id_institucion)
    JOIN dim_municipios
    USING(muni_id_ente)
GROUP BY
    1,2
ORDER BY
    1,2;
```

/PROMEDIO DE ESTUDIANTES CON DISCAPACIDAD COGNITIVA ÁREA
DE MATEMÁTICAS CLASIFICADOS POR ZONA Y MUNICIPIOS**/**

```
SELECT
    muni_nombre,
    ins_zona,
    ROUND(AVG(puntaje_matematicas),2) AS promedio
FROM
    dim_instituciones
    JOIN fact_saber5
    USING(id_institucion)
    JOIN dim_municipios
    USING(muni_id_ente)
    JOIN dim_estudiante
    USING(id_estudiante)
    WHERE estu_dissenso='S'
GROUP BY
    1,2
ORDER BY
    1,2;
```

/**PROMEDIO DE ESTUDIANTES CON DISCAPACIDAD COGNITIVA ÁREA DE MATEMÁTICAS CLASIFICADOS POR MUNICIPIOS/***

```
SELECT
    muni_nombre,
    ins_zona,
    ROUND(AVG(puntaje_matematicas),2) AS promedio
FROM
    dim_instituciones
    JOIN fact_saber5
    USING(id_institucion)
    JOIN dim_municipios
    USING(muni_id_ente)
    JOIN dim_estudiante
    USING(id_estudiante)
    WHERE estu_dissenso='S'
GROUP BY
    1,2
ORDER BY
    1,2;
```

/** PROMEDIO DE ESTUDIANTES CON DISCAPACIDAD COGNITIVA ÁREA DE CIENCIAS CLASIFICADOS POR MUNICIPIOS/***

```
SELECT
    muni_nombre,
    ins_zona,
    ROUND(AVG(puntaje_ciencias),2) AS promedio
FROM
    dim_instituciones
    JOIN fact_saber5
    USING(id_institucion)
    JOIN dim_municipios
    USING(muni_id_ente)
    JOIN dim_estudiante
    USING(id_estudiante)
    WHERE estu_dissenso='S'
GROUP BY
    1,2
ORDER BY
    1,2;
```

/**PROMEDIO DE ESTUDIANTES CON DISCAPACIDAD COGNITIVA ÁREA DE COMPETENCIAS CLASIFICADOS POR MUNICIPIOS**/

```
SELECT
  muni_nombre,
  ins_zona,
  ROUND(AVG(puntaje_competencias),2) AS promedio
FROM
  dim_instituciones
  JOIN fact_saber5
  USING(id_institucion)
  JOIN dim_municipios
  USING(muni_id_ente)
  JOIN dim_estudiante
  USING(id_estudiante)
  WHERE estu_dissenso='S'
GROUP BY
  1,2
ORDER BY
  1,2;
```


ANEXO 2. EXPOSICIÓN: ENCUENTRO NACIONAL DE ESTUDIANTES DE INGENIERÍA INFORMÁTICA Y AFINES.

Se realizó la exposición en la Institución Universitaria Colegio Mayor del Cauca, sobre la investigación realizada en el presente trabajo de grado.

Figura 50. Certificado del Encuentro nacional de estudiantes de ingeniería informática y afines.



Fuente: Encuentro Nacional de estudiantes de ingeniería y afines, realizado del 4 al 6 de octubre del 2017 en la ciudad de Popayán, Cauca.

ANEXO 3. POSTER: CACIED 2017

Se realizó la presentación a modo poster la investigación realizada en el presente trabajo de grado en la Institución Universitaria CESMAG en Pasto – Nariño.

Figura 51. Certificado del 3er congreso andino en computación, informática y educación.



Fuente: Congreso andino en computación, informática y educación, realizado del 1 al 3 de noviembre del 2017 en Pasto, Nariño.