DETECCIÓN DE PATRONES ASOCIADOS AL DESEMPEÑO **ACADÉMICO EN LAS PRUEBAS SABER 5°** DE LOS ESTUDIANTES DEL DEPARTAMENTO DE NARIÑO CON ÁRBOLES DE DECISIÓN

Silvio Ricardo Timarán Pereira
Arsenio Hidalgo Troya
Segundo Javier Caicedo Zambrano

Editorial
Universidad de Nariño

êditorial
Universidad de **Nariño**

DETECCIÓN DE PATRONES ASOCIADOS AL DESEMPEÑO **ACADÉMICO EN LAS PRUEBAS SABER 5°** DE LOS ESTUDIANTES DEL DEPARTAMENTO DE NARIÑO CON ÁRBOLES DE DECISIÓN

DETECCIÓN DE PATRONES ASOCIADOS AL DESEMPEÑO **ACADÉMICO EN LAS PRUEBAS SABER 5°** DE LOS ESTUDIANTES DEL DEPARTAMENTO DE NARIÑO CON ÁRBOLES DE DECISIÓN

Silvio Ricardo Timarán Pereira
Arsenio Hidalgo Troya
Segundo Javier Caicedo Zambrano

êditorial
Universidad de Nariño

Timarán Pereira, Silvio Ricardo
Detección de patrones asociados al desempeño académico en las pruebas saber 5° de los estudiantes del departamento de Nariño con árboles de decisión / Silvio Ricardo Timarán Pereira, Arsenio Hidalgo Troya, Segundo Javier Caicedo Zambrano-- San Juan de Pasto : Editorial Universidad de Nariño, 2025

168 páginas : ilustraciones, gráficas, tablas
Incluye referencias bibliográficas p. 152-162 y reseña de los autores p. 166
ISBN: 978-628-7771-66-6

1. Pruebas saber 5°--Evaluación educativa—Colombia 2. Pruebas saber 5°--Instituciones educativas básica primaria—Nariño (Colombia) 3. Sistema educativo colombiano—Estructura 4. Desempeño académico—Pruebas saber 5° 5. Competencias Lenguaje y matemáticas—Desempeño pruebas saber 5° 6. Variables socioeconómicas—Efectos—Desempeño pruebas saber 5° 7. Variables Institucionales—Efectos—desempeño pruebas saber 5° I. Hidalgo Troya, Arsenio II. Caicedo Zambrano, Segundo Javier.

371.260986158 T582 – SCDD-Ed. 22

Universidad de Nariño
FUNDADA EN 1904

Universidad de Nariño
ACREDITADA EN ALTA CALIDAD
RESOLUCIÓN MEN 000022 - ENERO 11 DE 2023

SECCIÓN DE BIBLIOTECA

DETECCIÓN DE PATRONES ASOCIADOS AL DESEMPEÑO ACADÉMICO EN LAS PRUEBAS SABER 5° DE LOS ESTUDIANTES DEL DEPARTAMENTO DE NARIÑO CON ÁRBOLES DE DECISIÓN

© Editorial Universidad de Nariño
© Silvio Ricardo Timarán Pereira
Arsenio Hidalgo Troya
Segundo Javier Caicedo Zambrano

ISBN: 978-628-7771-66-6

Corrector de estilo: Germán Chaves Jurado

Diseño y Diagramación: Alejandra Daniela Garzón Rodríguez

Fecha de publicación: Julio 2025
San Juan de Pasto - Nariño - Colombia

Prohibida la reproducción total o parcial, por cualquier medio o con cualquier propósito, sin la autorización escrita de su Autor o de la Editorial Universidad de Nariño.

AGRADECIMIENTOS

Al Sistema de Investigaciones de la Universidad de Nariño, por financiar esta investigación.

Tabla de contenido

INTRODUCCIÓN *10*

CAPÍTULO I / CONTEXTO DE LA PRUEBA Saber 5º *14*

Perspectiva Legal *14*

Estructura del Sistema Educativo Colombiano *17*

Pruebas Saber 5º *19*

Elaboración de las Pruebas *22*

Áreas y Preguntas *23*

Resumen de las competencias evaluadas en las Pruebas Saber 5º *42*

Modelos de Factores Asociados al Desempeño Escolar *43*

Modelo Ecológico *44*

Modelo de Eficacia Escolar de Scheerens *45*

Modelo Contexto, Insumos, Procesos y Productos(CIPP) *47*

Factores Asociados a las Pruebas Saber 3º, Saber 5º y Saber 9º *49*

Factores Asociados al Desempeño Académico en las Pruebas Saber *51*

Variables de Contexto Relacionadas con Características de los Estudiantes *55*

Variables de Contexto Relacionados con Características de las Instituciones Educativas *59*

Variables Relacionadas con la Dimensión Insumos *60*

Variables de Procesos Escolares *64*

Variables Relacionadas con Resultados Educativos *67*

Estudios Generales *68*

CAPÍTULO II / EFECTOS EN EL DESEMPEÑO ACADÉMICO DE LAS PRUEBAS SABER 5° EN EL DEPARTAMENTO DE NARIÑO..... 70

Características Socioeconómicas 72

Características Institucionales 74

Resultados de los Estudiantes Nariñenses en las Pruebas de Lenguaje y Matemáticas del Saber 5º 78

Efecto Localidad en las Pruebas de Lenguaje y Matemáticas en el departamento de Nariño.....79

Efecto Localidad en Lenguaje en el Departamento de Nariño... 79

Efecto Localidad en Matemáticas en el Departamento de Nariño ..81

Efecto de las Variables Socioeconómicas en el Desempeño de las Competencias de Lenguaje y Matemáticas en las Pruebas Saber 5º del Departamento de Nariño..... 83

Efecto de la prueba de Lenguaje 84

Efecto en la prueba de Matemáticas..... 86

Efecto de las Variables Institucionales en el Desempeño de las Competencias de Lenguaje y Matemáticas en las Pruebas Saber 5º del Departamento de Nariño..... 89

Efecto en la prueba de Lenguaje 89

Efecto en la prueba de Matemáticas.....90

CAPÍTULO III / MATERIALES Y MÉTODOS..... 92

Materiales..... 92

Metodología..... 102

Metodología CRISP-DM 103

Fase de Análisis del Problema *104*

Fase de Análisis de los Datos *106*

Fase de Preparación de los Datos *107*

Fase de Modelado *113*

Fase de Evaluación *115*

Fase de Explotación *115*

CAPÍTULO IV / RESULTADOS E INTERPRETACIÓN DE PATRONES 116

Definición del Modelo ***117***

Ajuste de Parámetros ***118***

Construcción de Modelos ***120***

Descubrimiento de Patrones de Desempeño Académico en la Competencia de Lenguaje en las Pruebas Saber 5º de los Estudiantes del Departamento de Nariño *121*

Descubrimiento de Patrones de Desempeño Académico en la Competencia de Matemáticas en las Pruebas Saber 5º de los Estudiantes del Departamento de Nariño *130*

CAPÍTULO V / CONCLUSIONES Y TRABAJOS FUTUROS 145

REFERENCIAS 152

Lista de figuras 163

Lista de tablas 164

Acerca de los autores 166

INTRODUCCIÓN

Las Pruebas Saber constituyen una estrategia que posibilita la evaluación de la calidad educativa en todos los establecimientos del país, tanto públicos como privados. Su propósito es monitorear el rendimiento académico, entendido como el desarrollo de competencias básicas de los estudiantes, y realizar un seguimiento a la calidad educativa proporcionada por las instituciones.

Hasta el Quinto Grado de Básica Primaria y Noveno Grado de Básica Secundaria, las pruebas Saber buscan valorar los avances en un determinado período, medir el impacto de los programas y constituirse en soporte para la propuesta de acciones de mejora. Estas pruebas evalúan competencias en Lenguaje, Matemáticas, Ciencias Naturales y Ciudadanas, estando alineadas con los Estándares Básicos de Competencias establecidos por el Ministerio de Educación Nacional (MEN).

Las Pruebas Saber 5° evalúan a los estudiantes de 5° Grado de Básica Primaria, el objetivo es contribuir al mejoramiento de la calidad de la Educación Colombiana mediante evaluaciones periódicas. Estos exámenes permiten identificar fortalezas y debilidades, facilitando la definición de planes de mejora por parte de los establecimientos educativos, Secretarías de Educación, el MEN y la sociedad en general. La periodicidad de las pruebas posibilita valorar avances en un período específico y establecer el impacto de los programas y de las acciones de mejora (Icfes, 2017).

Los resultados de evaluaciones nacionales e internacionales indican que Colombia tiene bajos logros académicos en todos los niveles de estudio (Posada y Mendoza, 2014). Esta situación es crítica, ya que, si persisten los bajos rendimientos académicos, se limitará el desarrollo del país y su crecimiento económico, tal como lo afirma la OECD (2016a).

Diversos estudios, como los de Torres et al. (2014), Martín (2015) y Gutiérrez (2014), han buscado identificar variables asociadas al rendimiento académico, centrándose en áreas específicas como Ciencias Naturales, Matemáticas o Lenguaje.

Un estudio del Icfes (2009) analizó factores asociados a las Pruebas Saber 5° y 9°, concluyendo que, a mayor nivel socioeconómico, los estudiantes obtienen un mejor desempeño. Además, encontró que, los alumnos de colegios privados suelen tener puntajes más altos y que las diferencias se acentúan con mejores condiciones socioeconómicas.

Asimismo, el Icfes ha realizado investigaciones, como el informe sobre factores asociados en las Pruebas Saber 5° y 9° (Icfes, 2011), utilizando técnicas estadísticas para visualizar elementos que inciden en el desempeño académico. También se ha estudiado la relación entre los factores asociados al rendimiento escolar a nivel de instituciones educativas, aulas de clase y estudiantes (Icfes, 2017).

Hasta ahora, los estudios sobre el desempeño académico en las Pruebas Saber 5°, se basan en análisis estadísticos, donde se consideran variables y relaciones primarias; sin embargo, dada la complejidad de las relaciones de las variables que inciden en el desempeño académico, se requiere un enfoque complementario, para lo cual, la minería de datos juega un papel importante, que surge como el siguiente paso evolutivo en el análisis de los datos y facilita el descubrimiento de diversos patrones y relaciones. La minería de datos es un proceso de detección de patrones válidos, novedosos, potencialmente útiles y fácilmente entendibles por el usuario a partir de los datos (Frawley, Piatetsky-Shapiro y Matheus, 1992).

El objetivo de la investigación que se presenta en este libro, fue detectar patrones asociados al desempeño académico de los estudiantes de instituciones de básica primaria del departamento de Nariño que presentaron las Pruebas Saber 5° en el año 2017.

Teniendo en cuenta lo expuesto por Frawley et al. (1992), se escogió el modelo de minería de datos clasificación basado en árboles de decisión, por ser el más adecuado para el problema objeto de esta investigación, por su simplicidad y facilidad para entender e interpretar los patrones descubiertos por parte de los investigadores. Hecho que no sucede con otros métodos como las redes neuronales artificiales que construyen modelos de caja negra, donde es difícil de entender la explicación de los resultados obtenidos.

Se seleccionaron los datos socioeconómicos, académicos e institucionales de los estudiantes del departamento de Nariño que presentaron las pruebas Saber 5º, de las bases de datos del Icfes. Se siguieron las fases de la metodología CRISP-DM para construir un conjunto de datos limpio y transformado y generar a partir de este conjunto, diferentes modelos de clasificación basados en árboles de decisión para descubrir patrones asociados al desempeño académico de estos estudiantes en las competencias de lenguaje, matemáticas y desempeño general en las pruebas Saber 5º del año 2017.

Esta investigación busca responder a la pregunta sobre los factores socioeconómicos, académicos e institucionales vinculados al desempeño académico en las Pruebas Saber 5° de los estudiantes de instituciones educativas de básica primaria del departamento de Nariño.

El documento se organiza en cinco capítulos. El Capítulo I, Contexto de las pruebas Saber, trata temas legales, estructura de las pruebas, el modelo de factores para la determinación de factores asociados al desempeño, y una revisión general de estudios relacionados con la presente investigación. El Capítulo II, Efectos en el Desempeño Académico, se analiza estadísticamente los resultados obtenidos en las Pruebas de Lenguaje y Matemáticas por los estudiantes del departamento de Nariño; asimismo, en aplicación del estadístico d de Cohen, se analiza a nivel departamental el efecto de la localidad, el efecto de las variables socioeconómicas y el

efecto de las variables institucionales en las competencias de lenguaje y matemáticas. El Capítulo III, Materiales y Métodos, describe de manera amplia la metodología CRISP-DM aplicada y los materiales utilizados. El Capítulo IV, Resultados e Interpretación de Patrones, define y evalúa el modelo aplicado, y presenta los patrones de Desempeño Académico en las Pruebas Saber 5° del departamento de Nariño, teniendo en cuenta el puntaje general y los puntajes en las pruebas de lenguaje y matemáticas. Finalmente, el Capítulo V, presenta conclusiones del estudio y recomendaciones para el desarrollo de trabajos futuros.

CAPÍTULO I
CONTEXTO DE LA PRUEBA Saber 5º

Perspectiva Legal

La Constitución Política Nacional de Colombia, establece que, la Educación es un derecho y un servicio público, que tiene una función social, a través de la cual, se busca el acceso al conocimiento, a la ciencia, a la técnica y a los demás bienes y valores de la cultura (Decreto 1400, 1970, Art. 67) por lo cual, corresponde al Estado, regular y ejercer la inspección y vigilancia de la Educación, en la vía de garantizar su calidad y el cumplimiento de sus objetivos.

Con la Ley 115 de 1994, se dispuso la creación de un Sistema Nacional de Evaluación de la Educación, a cargo del Ministerio de Educación Nacional (MEN) y del Instituto Colombiano para el Fomento de la Educación Superior (Icfes), hoy, Instituto Colombiano para la Evaluación de la Educación, donde se sentaron las bases para el diseño e implementación de un nuevo sistema de evaluación: la interna o de aula y la externa.

En cuanto a la evaluación interna, Naranjo (2011) afirma que, "Lo que debe evaluar la evaluación interna, son desarrollos de pensamiento, de actitudes, de valores y niveles de aprendizaje del estudiante, para lo cual, los contenidos, son medios para lograr esos aprendizajes y no fines en sí mismos" (p.52). Por su parte, la Ley 1324 de 2009[1], confiere al Icfes

1 *Esta Ley establece que: "[...] La estructura de los exámenes, deberá mantenerse por períodos no menores a 12 años, sin perjuicio de que se incluyan áreas o estudios particulares, que no alteren su comparabilidad en el tiempo. La presentación de los "exámenes de Estado", es requisito para ingresar a los programas de pregrado y obtener el título respectivo [...]." (Artículo 7º).*

(2016b)[2] la misión de evaluar mediante exámenes externos estandarizados la formación que se ofrece en el servicio educativo en los diversos niveles de la educación; dichos exámenes, están estandarizados en cuanto, a que, las condiciones de aplicación y procesamiento de los resultados son uniformes. También establece que, el MEN es quien define lo que debe evaluarse en estos exámenes (Icfes, 2014a).

Ahora, en cuanto a la evaluación interna, de acuerdo con el Decreto 1290 del MEN (2009), se caracteriza por ser formativa, motivadora, orientadora, pero nunca sancionatoria; debe ser transparente, continua y procesual (Rodríguez y Ahumada, 2017).

Desde el año 2002, la evaluación tiene un cambio de perspectiva y pasa de un enfoque de conocimientos a una evaluación de competencias[3]. "No obstante, es importante aclarar, que competencias y conocimientos no son elementos independientes, pues los conocimientos son necesarios para el desarrollo de las competencias" (Ayala, 2015). Los autores Ahumada *et al*., (2014) entienden las competencias, como capacidades complejas, que permiten a un sujeto, actuar de manera pertinente y adecuada con la situación en que se halla, disponiendo sus saberes, valores y sentires, expresados a través de una acción.

---

2 *De acuerdo con el artículo 12 Ley 1324 de 2009, el ICFES, es una empresa estatal de carácter social, del sector educativo nacional. Es una Entidad pública descentralizada, del orden nacional, de naturaleza especial, con personería jurídica, autotomía administrativa y patrimonio propio, vinculada al Ministerio de Educación Nacional, que tiene por objeto, ofrecer el servicio de evaluación de la educación en todos sus niveles y adelantar investigación sobre los factores que inciden en la calidad educativa, con la finalidad de ofrecer información para mejorar la actualidad de la educación (ICFES, 2016b).*

3 *La noción de competencia, según Vasco (2003), se define como: "Conjunto de conocimientos, habilidades, actitudes, comprensiones y disposiciones cognitivas, metacognitivas, socioafectivas, comunicativas y psicomotoras, apropiadamente relacionadas entre sí, para facilitar el desempeño flexible, eficaz y con sentido de una actividad o de cierto tipo de tareas, en contextos relativamente nuevos y retadores".*

Las competencias se evalúan con base en los *Estándares Básicos de Competencias* que establecen lo que deben Saber y Saber hacer los estudiantes, los cuales se constituyen en "referentes que permiten evaluar los niveles de desarrollo de las competencias que van alcanzando los y las estudiantes en el transcurrir de su vida escolar" (Rodríguez y Ahumada, 2017).

En la evaluación externa, reconoce la importancia de las evaluaciones censales y muestrales en la toma de decisiones. Según el Informe de Gestión y Rendición de Cuentas del MEN (2014), los resultados de las Pruebas Saber aplicadas en 2009, evaluaron el desempeño de los estudiantes de 5º y se constituyeron en la línea base para las aplicaciones más recientes. En 2009, las Pruebas evidenciaron que un alto porcentaje de estudiantes se ubicaba en los niveles insuficiente y mínimo de desempeño; particularmente, en Grado Quinto; los porcentajes en dichos niveles alcanzaron el 67% en Matemáticas y en Grado Noveno, el 73% (Rodríguez y Ahumada, 2017).

Con el fin de identificar las principales diferencias y brechas con los estándares internacionales de calidad educativa, Colombia ha participado en diversas pruebas internacionales, así: Estudio de Tendencias Internacionales y Ciencias -TIMSS; el Segundo Estudio Regional Comparativo y Explicativo -SERCE; el Estudio Internacional de Educación Cívica y Ciudadanía -ICCS y el Programa Internacional de Evaluación de Estudiantes -PISA. En las Pruebas PISA, el compromiso de Colombia, era mantener su participación de forma que se pueda monitorear los avances en el tiempo (Rodríguez y Ahumada, 2017, p. 277).

Teniendo en cuenta el Plan Nacional Decenal de Educación 2006-2016, el Icfes avanzó en la alineación del Sistema Nacional de Evaluación Externa Estandarizada (SNEE) mediante la reestructuración de los exámenes, así: en 2009, crea un nuevo diseño de Saber 3°, 5° y 9°; en 2010 rediseñó la Prueba Saber Pro; en 2014, implementó cambios en la prueba Saber 11°. Dicha alineación, posibilita la comparación de resultados entre los niveles educativos (Icfes, 2014b).

La Tabla 1, resume las competencias que se evalúan en cada Prueba de 3°, 5°, 9°, 11° y en el Saber Pro, donde se puede apreciar la alineación de las mismas.

Tabla 1. Competencias evaluadas en cada Prueba Saber

3°	5°	9°	11°	Pro
Lenguaje	Lenguaje	Lenguaje	Lectura Crítica	Lectura Crítica
Matemáticas	Matemáticas	Matemáticas	Matemáticas	Razonamiento Cuantitativo
	Competencias Ciudadanas	Competencias Ciudadanas	Sociales y Ciudadanas	Competencias Ciudadanas
	Ciencias Naturales	Ciencias Naturales	Ciencias Naturales	Inglés
		Inglés	Inglés	Comunicación Escrita
				Módulos Específicos

Fuente: elaboración propia con datos de ***Icfes*** (s.f).

Estructura del Sistema Educativo Colombiano

El Sistema Educativo Colombiano, está organizado en cuatro niveles:

Educación Preescolar: incluye tres años y corresponde a la educación de niños menores de seis años; para las instituciones oficiales es obligatorio el Grado de Transición.

Educación Básica: incluye los ciclos de Básica Primaria, que corresponde a los Grados de Primero a Quinto y Básica Secundaria, los Grados de Sexto a Noveno.

Educación Media: corresponde a los Grados Diez y Once.

Educación Superior: se clasifica en Educación Técnica, Tecnológica y Universitaria; la duración oscila entre 1 y 7 años.

En el Sistema Educativo Colombiano, son obligatorios los Grados de Transición y Noveno, para un total de 10 años.

Los esfuerzos del país, por la oferta de una Educación que se adapte a las condiciones y características de la población rural, han tenido efectos positivos en alfabetismo, escolaridad promedio y asistencia escolar. Sin embargo, se mantiene la brecha educativa, en relación con la zona urbana y rural, por lo cual, es necesario que se continúe con estrategias educativas que se adapten a las condiciones rurales de las regiones del país (Fedesarrollo, 2014).

Las dificultades de las comunidades indígenas, en cuanto al acceso y la permanencia, se relacionan, con deficiencias en la oferta en los territorios y con la falta de pertinencia de la educación. A pesar de que la Ley 115 de 1994 establece la necesidad de ofrecer educación diferencial a los grupos étnicos, recientemente se viene trabajando en la formulación de proyectos etno-educativos, con la participación de las comunidades, en los que se establecen los contenidos, considerando la autonomía, lengua y plan de vida de cada pueblo indígena (Fedesarrollo, 2014).

Inicialmente, la Prueba Saber 11 era el único examen que se aplicaba para evaluar los aprendizajes de los estudiantes, el cual, hoy se aplica anualmente a todos los estudiantes de Grado 11 y se realiza en las áreas de Lenguaje, Matemáticas, Ciencias Sociales, Química, Física, Filosofía, Biología e Inglés. A partir de los años 2002 y 2003, también se aplican Pruebas Saber a los estudiantes de los Grados Quinto y Noveno y se enfocan en las áreas de Lenguaje, Matemáticas y Ciencias y se realizan cada tres años (Fedesarrollo, 2014).

Pruebas Saber 5º

Las Pruebas Saber 5° buscan contribuir al mejoramiento de la calidad de la educación en el país; su aplicación es periódica. Se evalúan las competencias básicas de los estudiantes y se analizan los factores que inciden en sus logros. Las pruebas son censales y muestrales. En la aplicación muestral, a nivel nacional, se selecciona un grupo representativo de estudiantes de establecimientos educativos oficiales y privados, urbanos y rurales; la prueba censal, se aplica a todos los estudiantes que se encuentren cursando dicho grado (MEN, 2017).

Entre los objetivos de la política educativa del MEN, está "[...] lograr una Educación de calidad, que forme mejores seres humanos, ciudadanos con valores éticos, competentes, respetuosos de lo público, que ejercen los derechos humanos, cumplen con sus deberes y conviven en paz. Una Educación que genere oportunidades legítimas de progreso y prosperidad para ellos y para el país" (MEN, 2017).

En este sentido, el MEN (2017):

Está implementando estrategias de mejoramiento de la calidad y pertinencia de la Educación, orientadas al fortalecimiento de las competencias básicas y ciudadanas del sistema nacional de evaluación, del sistema de aseguramiento de la calidad, como también, al desarrollo profesional de directivos y docentes del magisterio colombiano (p. 8).

A partir del año 2015, las Pruebas Saber, contienen dos de los componentes evaluativos del Índice Sintético de Calidad Educativa (ISCE), métrica que utiliza el MEN para el análisis del desempeño académico de los estudiantes de instituciones educativas de educación básica y media del país.

En todo caso, la Prueba Saber 5° busca "Contribuir al mejoramiento de la calidad de la Educación Colombiana, mediante la realización de pruebas periódicas, en las que se evalúan las competencias básicas de los estudiantes y se analizan los factores que inciden en sus logros". (MEN, 2017).

Así pues, los resultados contribuyen a que los establecimientos educativos, las secretarías de educación, el MEN y la sociedad en general, conozcan las fortalezas y aspectos por mejorar, en la vía del aseguramiento de la calidad educativa; a la vez, constituyen un insumo para la formulación de sendos planes de mejoramiento.

En la Tabla 2 y la Tabla 3, se resume, grosso modo, los cambios que han tenido las Pruebas Saber 3°, 5° y 9° desde 1991 hasta el 2006. Se destaca que, inicialmente, se aplicaron a muestras de estudiantes, pero desde 2002 se procede a una aplicación a todo el universo de estudiantes.

Las Pruebas Saber, evalúan las competencias que desarrollan los estudiantes entre 4° y 5° grados, en relación con los estándares básicos de competencias establecidos por el MEN, que constituyen el referente para establecer "[...] qué tanto los estudiantes y el sistema educativo en su conjunto, están cumpliendo con las expectativas de calidad educativa" (MEN, 2017).

Las Pruebas Saber, no cubren todos los contenidos ni los estándares que se definen para cada área de evaluación; además, algunos logros únicamente se pueden valorar en el desarrollo de las actividades escolares. Sin embargo, son un indicador sobre el avance en el proceso de formación en Educación Básica y resultan un insumo para establecer niveles de desempeño y para la identificación de las fortalezas y los aspectos por mejorar de los estudiantes, como también, para la formulación de estrategias y planes de mejoramiento.

*Tabla 2. **Antecedentes de las Pruebas Saber** 3°, 5° y 9°: **Pruebas Muestrales de Lenguaje, Matemáticas y Ciencias***

Año/Descripción	Pruebas/Aspectos destacados
1991	Muestra representativa para algunas ciudades de 13 departamentos del país
1992-1995	Muestra a nivel nacional y por regiones Tipo de resultados que publica: puntaje promedio, desviación estándar
1997-1999	Muestra representativa a nivel nacional y por departamentos

Fuente: elaboración propia con datos de ***Icfes*** (s.f).

*Tabla 3. **Antecedentes de las Pruebas Saber 3°, 5° y 9°: Pruebas Censales de Lenguaje, Matemáticas y Ciencias Naturales, Ciencias Sociales y Competencias Ciudadanas***

Año/Descripción	Aspectos destacados
2002-2003	• Aplicación cada tres años
2005-2006	• Se reportan datos a nivel de institución, municipio, departamento y país. • Resultado: niveles de logro, puntaje promedio y desviación estándar, componentes del área, porcentaje de estudiantes por opción de respuesta

Fuente: elaboración propia con datos de ***Icfes*** (s.f).

Elaboración de las Pruebas

Las pruebas son diseñadas por el Icfes, con apoyo de equipos de expertos de cada área; para el efecto, se aplica el Modelo Basado en Evidencias (MBE), en el cual, las especificaciones se formalizan, iniciando con las afirmaciones sobre las competencias que debe poseer un estudiante en su desempeño en el módulo. Posteriormente, se describen las evidencias que deben sustentar cada una de las afirmaciones y, finalmente, se establecen las tareas que debe realizar el estudiante, como evidencia de las afirmaciones (MEN, 2017).

La Tabla 4 presenta, de manera resumida, las fases que se desarrollan para realizar la evaluación de la Educación del Sistema Educativo Colombiano, en las cuales, participan el MEN, el Icfes, las secretarías de educación y las instituciones educativas.

*Tabla 4. **Fases de la Evaluación***

Institución	**Actividades**
MEN	• Define, políticas, propósitos y usos de las evaluaciones • Define, los referentes de lo que se desea evaluar en consulta con los grupos de interés • Realiza, seguimiento a estrategias y planes de mejoramiento

Icfes	• Diseña, construye y aplica, las evaluaciones • Analiza y divulga los resultados • Identifica los aspectos críticos
Secretaría de Educación y Establecimientos Educativos	• Diseñan, implementan y coordinan estrategias y planes de mejoramiento

Fuente: elaboración propia con datos de ***Icfes*** (s.f).

Áreas y Preguntas

La Prueba Saber 5° evalúa las áreas de Matemáticas, Lenguaje, Ciencias Naturales y Competencias Ciudadanas; contiene preguntas de selección múltiple, con única respuesta, que incluyen: un enunciado relacionado con una situación, contexto y texto; la formulación de una tarea de evaluación, donde se establece lo que debe realizar el estudiante; y cuatro opciones de respuesta, de las cuales, sólo una es correcta y válida según la tarea que se plantea. Desde el 2012, la prueba se aplica anualmente y siempre evalúa Competencias de Lenguaje y Matemáticas, y no siempre Ciencias Naturales y Competencias Ciudadanas.

Durante la Prueba Cognitiva, también se entrega un cuadernillo, que incluye un cuestionario de contexto, mediante el cual, se recoge información de factores asociados al desempeño académico; se obtiene información, sobre indicadores que tienen relación con los procesos de enseñanza y de aprendizaje, y que se pueden utilizar para explicar los desempeños de los estudiantes en las pruebas. Concretamente, se recoge datos del núcleo familiar: composición, estatus laboral y educativo; condiciones del hogar, dotación de bienes dentro de la vivienda, situación económica, estrato socioeconómico, disponibilidad de internet, servicio de televisión por cable y dedicación familiar a entretenimiento (MEN, 2017).

Prueba de Lenguaje. La prueba asume procesos de comprensión y de producción; se organizan en torno a cinco factores: (a) producción textual; (b) comprensión e interpretación textual; (c) literatura, que supone un abordaje de la perspectiva estética del lenguaje; (d) medios de comunicación y otros sistemas simbólicos; y (e) ética de la comunicación, el cual supone, como eje transversal, la responsabilidad de reflexionar sobre el tratamiento de la información, el reconocimiento de discursos y estrategias de comunicación situadas en los fenómenos y procesos sociales (MEN, 2017).

> Los Procesos de Comprensión, hacen referencia a la búsqueda y reconstrucción del sentido y los significados presentes en diversos tipos de textos: literarios, informativos, descriptivos, avisos, tablas, gráficos, entre otros; y otras formas de comunicación no verbal, como gestos, música y expresiones artísticas en general. Los Procesos de Producción se relacionan con generación de significados, tanto para expresarse como para transmitir información o interactuar con los demás, atendiendo a las reglas que regulan la comunicación en cada uno de los sistemas de significación verbal o no verbales. Se busca que, el estudiante ponga en juego los conocimientos sobre el uso y funcionamiento de la lengua en situaciones comunicativas particulares, de modo que, responda a las necesidades de comunicación: exponer, narrar, argumentar, entre otras; y también, dé respuesta a las exigencias particulares del contexto de comunicación, orientado por procedimientos sistemáticos para producir textos coherentes y cohesionados (MEN, 2017, p. 17).

> La prueba evalúa dos competencias: la Comunicativa-Lectora y la Comunicativa-Escritora.

> La *Comunicativa-Lectora*, incluye la comprensión, el uso y la reflexión sobre las informaciones contenidas en diversos tipos de textos; implica una relación dinámica entre estos y el lector. Propone una reflexión, en torno a qué dice el texto (contenidos conceptuales e ideológicos), cómo lo dice (organización), para qué lo dice y por qué lo

> dice (pragmática), cuándo lo dice y quién lo dice. Las preguntas de la prueba, se orientan a establecer la capacidad de los estudiantes para realizar lecturas literales, inferenciales o críticas (MEN, 2017, p. 17).
>
> La *Competencia Comunicativa-Escritora*, de acuerdo a los estándares, corresponde a la producción de textos escritos, de manera que atiendan a los siguientes requerimientos: (a) responder a las necesidades comunicativas, es decir, se requiere relatar, informar, exponer, solicitar o argumentar sobre un determinado tema; (b) cumplir procedimientos sistemáticos para su elaboración; y (c) utilizar los conocimientos de la persona que escribe (MEN, 2017, p. 18).
>
> En relación con esta competencia, se evalúa el proceso de escritura y no la escritura en sí misma; es decir, no se pide la elaboración de textos escritos, sino que, las preguntas indagan sobre los tipos de textos que utilizarían para lograr un determinado propósito o finalidad comunicativa, la forma cómo los organizarían para expresar un mensaje o una idea, y aspectos relacionados con el uso adecuado de las palabras y frases para producir textos con sentido (MEN, 2017, p. 18).

Por otra parte, la Prueba de Lenguaje evalúa tres componentes: semántico, sintáctico y pragmático. El primero, hace referencia al sentido del texto en términos de su significado, indaga por el qué se dice en el texto; el segundo, se refiere a la organización del texto en términos de su coherencia y cohesión, indaga por el cómo se dice; y el tercero, tiene que ver con, para qué se dice, en función de la situación de comunicación (MEN, 2017).

La Tabla 5 y la Tabla 6, contienen las afirmaciones elaboradas para cada componente evaluado en la prueba y para cada competencia.

*Tabla 5. **Competencia Comunicativa Lectora – Grados 4° y 5°***

Componente	**Afirmación**
Semántico	Recupera información explícita contenida en el texto
	Recupera información implícita contenida en el texto
	Relaciona textos entre sí y recurre a Saberes previos para ampliar referentes e ideas
Sintáctico	Identifica la estructura explícita del texto
	Recupera información implícita de la organización, la estructura y los componentes de los textos
	Evalúa estrategias explícitas o implícitas de organización, estructura y componentes de los textos
Pragmático	Reconoce información explícita sobre los propósitos del texto
	Reconoce elementos implícitos sobre los propósitos del texto
	Analiza información explícita o implícita sobre los propósitos del texto

Fuente: Icfes (2014b).

Tabla 6. Competencia Comunicativa Escritora – Grados 4° y 5°

Componente	Afirmación
Semántico	Prevé, temas, contenidos, ideas o enunciados, para producir textos que respondan a diversas necesidades comunicativas
	Realiza consultas con base en las características del tema y el propósito del texto
	Comprende los elementos formales, que regulan el desarrollo de un texto, teniendo en cuenta lo que quiere comunicarse
Sintáctico	Prevé el plan, para organizar las ideas y para definir el tipo de texto pertinente, de acuerdo con lo que quiere comunicar
	Conoce la organización, que un texto debe tener para logar coherencia y cohesión
	Conoce los elementos formales de la lengua y de la gramática, para lograr la coherencia y la cohesión del texto, en una situación de comunicación particular
Pragmático	Prevé el propósito o las intenciones que un texto debe cumplir, para atender las necesidades de la comunicación
	Utiliza las estrategias discursivas pertinentes y adecuadas de acuerdo con el propósito de la comunicación que debe tener un texto
	Utiliza los elementos formales de las estrategias discursivas, con el fin de adecuar el texto a la situación de comunicación

Fuente: Icfes (2014b).

Prueba de Matemáticas. La evaluación se refiere a la capacidad que tiene el individuo para implementar conceptos y estructuras matemáticas en un contexto determinado y las estrategias que utiliza para la interpretación. Se busca poner en evidencia, los significados que el estudiante ha logrado construir y que aplica en la resolución de diversas situaciones problema; en tal sentido, se evalúa el significado de los conceptos matemáticos y su puesta en práctica mediante la matematización, misma que exige simbolizar, formular, cuantificar, validar, representar, generalizar, entre otras acciones (MEN, 2017).

La prueba de matemáticas, evalúa competencias matemáticas de comunicación, modelación, razonamiento, planteamiento y resolución de problemas, elaboración, comparación y ejercitación de procedimientos, que son establecidas por los lineamientos curriculares del MEN. Para efectos de la prueba, dichas competencias se reagrupan, así: el razonamiento y la argumentación, la comunicación, la representación y la modelación, y el planteamiento y resolución de problemas, las cuales incluyen la elaboración, comparación y ejercitación de procedimientos (Ibidem), competencias que se describan en seguida.

> El *Razonamiento y la Argumentación*, se relacionan con la capacidad para dar cuenta del cómo y del porqué de los caminos que se siguen para llegar a conclusiones, justificar estrategias y procedimientos puestos en acción, en el tratamiento de situaciones problema, formular hipótesis, proponer opiniones e ideas, explorar ejemplos y contraejemplos, probar y estructurar argumentos, generalizar propiedades y relaciones, identificar patrones y expresarlos matemáticamente, y plantear preguntas, reconocer distintos tipos de razonamiento y distinguir y evaluar cadenas de argumentos (MEN 2017, p. 33).

> La *Comunicación, Representación y Modelación*, se refieren a la capacidad del estudiante para expresar ideas, interpretar, utilizar

diferentes tipos de representación, describir relaciones matemáticas, describir situaciones o problemas utilizando el lenguaje escrito, concreto, pictórico, gráfico y algebraico, manipular expresiones que contengan símbolos y fórmulas, utilizar variables y describir cadenas de argumentos orales y escritas, traducir, interpretar y distinguir entre diferentes tipos de representaciones, interpretar lenguaje formal y simbólico, así como, traducir de lenguaje natural al simbólico formal y viceversa, que se resume en decodificar de manera entendible aquello expresado matemáticamente en palabras sencillas y manejables por el estudiante. (MEN 2017, p. 33).

Planteamiento y Resolución de Problemas, refieren a la capacidad para formular problemas a partir de situaciones dentro y fuera de las matemáticas; desarrollar, aplicar diferentes estrategias y justificar la elección de métodos e instrumentos para la solución de problemas; justificar la pertinencia de un cálculo exacto o aproximado en la solución de un problema y lo razonable o no de una respuesta obtenida; verificar e interpretar resultados a la luz del problema original; y generalizar soluciones y estrategias para dar solución a nuevas situaciones problema (MEN 2017, p. 34).

Para efectos de la estructuración de la prueba, los cinco pensamientos establecidos en los lineamientos curriculares y en los estándares básicos de competencias del MEN, se reorganizan en tres componentes: el numérico-variacional, el geométrico-métrico y el aleatorio, donde cada uno incluye aspectos, operaciones y relaciones entre cantidades, medidas y figuras, y, experimentos y probabilidad de eventos.

El Componente Numérico-Variacional, corresponde a aspectos asociados a los números y a la numeración, su significado y la estructura del sistema de numeración; las operaciones, sus propiedades, su efecto y las relaciones entre ellas. El reconocimiento

de regularidades y patrones que implica establecer, cuál es el cambio constante de una serie de valores o cómo estos se comportan, la identificación de variables, la descripción de fenómenos de cambio y dependencia; conceptos y procedimientos asociados, a la variación directa, a la proporcionalidad, a la variación lineal en contextos aritméticos y geométricos, el lenguaje simbólico (algebraico) articulado entre convenciones y esquemas o representaciones introductorias al manejo de variables, a la variación inversa y el concepto de función (MEN 2017, p. 34).

El *Componente Geométrico - Métrico*, está relacionado con la construcción y manipulación de representaciones de los objetos del espacio, las relaciones entre ellos y sus transformaciones; más específicamente, con la comprensión del espacio, el análisis abstracto de figuras y formas en el plano y en el espacio, a través de la observación de patrones y regularidades, el razonamiento geométrico y la solución de problemas de medición, la descripción y estimación de magnitudes [...], transformaciones de figuras representadas en el plano o en el espacio, la selección de unidades de medida, de patrones y de instrumentos, el uso de unidades, los conceptos de perímetro, área y volumen (MEN, 2017, p. 34).

Finalmente, el *Componente Aleatorio*, corresponde a la representación, lectura e interpretación de datos en contexto; el análisis de diversas formas de representación de información numérica, el análisis cualitativo de regularidades, de tendencias y la formulación de inferencias y argumentos, utilizando medidas de tendencia central y de dispersión y por el reconocimiento, descripción y análisis de eventos aleatorios (MEN 2017, p. 34).

Las tablas: Tabla 7, Tabla 8 y la Tabla 9, contienen las afirmaciones elaboradas para cada componente evaluado en la prueba y para cada competencia.

Tabla 7. Competencia: Comunicación, Representación y Modelación - 4° y 5°

Componente	Afirmación
Numérico Variacional	Reconoce significados del número en diferentes contextos (medición, conteo, comparación, codificación, localización, entre otros).
	Reconoce diferentes representaciones de un mismo número
	Describe e interpreta propiedades y relaciones de los números y sus operaciones
	Traduce relaciones numéricas expresadas gráfica y simbólicamente
Geométrico-Métrico	Establece relaciones entre los atributos mensurables de un objeto o evento y sus respectivas magnitudes
	Identifica unidades tanto estandarizadas como no convencionales apropiadas para diferentes mediciones y establece relaciones entre ellas
	Utiliza sistemas de coordenadas para especificar localizaciones
Aleatorio	Clasifica y organiza la presentación de datos
	Interpreta cualitativamente datos relativos a situaciones del entorno escolar
	Representa un conjunto de datos e interpreta representaciones gráficas de un conjunto de datos
	Hace traducciones entre diferentes representaciones
	Expresa el grado de probabilidad de un suceso

Fuente: Icfes (2014b).

Tabla 8. Competencia: Razonamiento y Argumentación - 4° y 5° Grados

Componente	**Afirmación**
Numérico Variacional	Reconoce patrones numéricos
	Justifica propiedades y relaciones numéricas utilizando ejemplos y contraejemplos
	Reconoce y genera equivalencias entre expresiones numéricas
	Analiza relaciones de dependencia en diferentes situaciones
	Utiliza y justifica propiedades (aditiva y posicional del sistema de numeración decimal)
Geométrico-Métrico	Compara y clasifica objetos tridimensionales y figuras bidimensionales de acuerdo con sus componentes
	Reconoce nociones de paralelismo y perpendicularidad en distintos contextos
	Hace conjeturas y verifica los resultados de aplicar transformaciones a figuras en el plano
	Describe y argumenta acerca del perímetro y del área de un conjunto de figuras planas cuando una de las magnitudes se fija
	Relaciona objetos tridimensionales y sus propiedades con sus respectivos desarrollos planos
	Construye y descompone figuras planas y sólidos a partir de condiciones dadas.
	Identifica y justifica relaciones de semejanza y congruencia entre figuras.

Aleatorio	Compara datos presentados en diferentes representaciones Hace arreglos condicionados o no condicionados.
	Hace conjeturas acerca de la posibilidad de ocurrencia de eventos.

Fuente: Icfes (2014b).

Tabla 9. Competencia: Planteamiento y Resolución de Problemas - 4° y 5° Grados

Componente	**Afirmación**
Numérico Variacional	Resuelve y formula problemas aditivos de transformación, comparación, combinación e igualación
	Resuelve y formula problemas multiplicativos de adición repetida, factor multiplicante, razón y producto cartesiano
	Resuelve y formula problemas de proporcionalidad directa e inversa
	Resuelve y formula problemas que requieren el uso de la fracción como parte de un todo, como cociente y como razón
Geométrico-Métrico	Utiliza diferentes procedimientos de cálculo para hallar la medida de superficies y volúmenes
	Reconoce el uso de las magnitudes y de las dimensiones de las unidades respectivas en situaciones aditivas y multiplicativas
	Utiliza relaciones y propiedades geométricas para resolver problemas de medición
	Utiliza y construye modelos geométricos para solucionar problemas

Aleatorio	Resuelve problemas que requieren representar datos relativos al entorno utilizando una o diferentes representaciones
	Resuelve problemas que requieren encontrar y/o dar significado al promedio de un conjunto de datos
	Resuelve situaciones que requieren calcular la posibilidad o imposibilidad de ocurrencia de eventos

Fuente: Icfes (2014b).

Prueba de Ciencias Naturales. Busca establecer y diferenciar las competencias de los estudiantes, en sus conocimientos básicos de las ciencias naturales, en comprensión y resolución de problemas; además, evalúa la comprensión de los estudiantes sobre las particularidades y los alcances del conocimiento científico, como también, la capacidad para diferenciar este conocimiento de otros Saberes (Icfes, 2014b).

La prueba reconoce la capacidad que tienen los estudiantes para lo siguiente: **a)** establecer relaciones significativas entre conceptos y en los conceptos propios de las Ciencias Naturales, con el fin de interpretarlos y actuar socialmente, de manera eficiente, honesta y ética; **b)** valorar los hechos y la calidad de las evidencias, que permiten interpretar y abordar una situación problema rigurosamente; **c)** formular preguntas y plantear problemas válidos; **d)** producir explicaciones en el ámbito de las ciencias naturales; **e)** construir diversas alternativas de solución o formas de actuar, frente a una situación problema y seleccionar con racionalidad la más adecuada; **f)** potenciar en la enseñanza de las ciencias en el país, el desarrollo de la capacidad para tomar decisiones utilizando conocimiento y la creatividad, con la cual, se fomenta la capacidad para hacer innovaciones, producir nuevas explicaciones y para contribuir a la transformación real de su entorno (Icfes, 2016b).

Concretamente, en esta prueba, se evalúan las siguientes competencias (Icfes, 2014b):

Uso Comprensivo del Conocimiento Científico: corresponde a la capacidad para comprender y utilizar conceptos, teorías y modelos, en la solución de problemas, a partir del conocimiento adquirido; se busca que, el estudiante relacione conceptos y conocimientos adquiridos con fenómenos que se observan con frecuencia; superar la repetición de conceptos, pasando a un uso comprensivo de los mismos.

Explicación de Fenómenos: refiere a la capacidad para construir explicaciones y comprender argumentos y modelos que den razón de fenómenos; tiene relación con la forma en que los estudiantes construyen sus explicaciones en el contexto de la ciencia escolar. Esta competencia, fomenta una actitud crítica y analítica, que le permite establecer la validez o coherencia de una afirmación o un argumento, con lo cual, puede presentar explicaciones de un mismo fenómeno, utilizando representaciones conceptuales pertinentes con diversos grados de complejidad (Icfes, 2009b, p. 21).

Indagación: corresponde a la capacidad para formular preguntas y procedimientos adecuados para buscar, seleccionar, organizar e interpretar información relevante con el fin de dar respuesta a dichas preguntas; el Proceso de Indagación en ciencias, implica observar detenidamente una situación, formular preguntas, buscar relaciones causa/efecto, recurrir a fuentes de información, hacer predicciones, plantear experimentos, identificar variables, realizar mediciones, y, organizar y analizar resultados (Icfes, 2009b, p. 21).

La prueba propone preguntas relacionadas con situaciones de la vida diaria, con el fin de estimular la necesidad de observar el medio y las situaciones del día a día, y de preguntar por los fenómenos desde la

perspectiva de las Ciencias Naturales. En este sentido y, en coherencia con los estándares básicos de competencias, las preguntas se inscriben en alguno de los siguientes componentes (Icfes, 2014b):

> ***Entorno Vivo:*** trata temas relacionados con los seres vivos y sus interacciones. Se centra en el organismo para entender sus procesos internos y sus relaciones con los medios físico y biótico; también, aborda los siguientes temas: estructura y función, homeóstasis, herencia y reproducción, ecología, evolución, diversidad y similitud. Por su parte, la salud se entiende, como el respeto y cuidado del cuerpo; también forma parte de este componente y de los componentes de Ciencia, Tecnología y Sociedad (CTS) (MEN, 2016, p. 50).
>
> ***Entorno Físico:*** se relaciona con la comprensión de conceptos, principios y teorías, a partir de los cuales, el estudiante describe y explica el mundo físico con el que interactúa. Se incluye el universo, haciendo énfasis en el sistema solar y la Tierra como planeta, y la materia y sus propiedades, estructura y transformaciones; también se incluyen nociones o conceptos como mezclas, combinaciones, reacciones químicas, energía, movimiento, fuerza, tiempo, espacio y sistemas de medición y nomenclatura (MEN, 2016, p. 50).
>
> ***Ciencia, Tecnología y Sociedad (CTS):*** se relaciona, con la independencia de criterio de los estudiantes, basada en conocimientos y evidencias, y un sentido de responsabilidad crítica, hacia el modo, cómo la ciencia y la tecnología pueden afectar sus vidas, las de sus comunidades y las del mundo en general. Se busca un mayor acercamiento entre la ciencia y el mundo del estudiante, fomentando una comprensión amplia del significado social de los conocimientos científicos y del desarrollo tecnológico (MEN, 2016, p. 50).

Las tablas que siguen, contienen las afirmaciones elaboradas para cada componente evaluado en la prueba y para cada competencia (Tabla 10, Tabla 11, Tabla 12).

Tabla 10. Competencia: Uso del Conocimiento Científico - 1° a 5° Grados

Componente	Afirmación
Entorno Vivo	Comprende que los seres vivos dependen del funcionamiento e interacción de sus partes
	Comprende que los seres vivos atraviesan diferentes etapas durante su ciclo de vida
	Comprende que existen relaciones entre los seres vivos y el entorno y que estos dependen de aquellas
Entorno Físico	Comprende que existe una gran diversidad de materiales que se pueden diferenciar a partir de sus propiedades
	Comprende que existen diversas fuentes y formas de energía y que esta se transforma continuamente
	Comprende la estructura básica y el funcionamiento de los circuitos eléctricos
	Reconoce los principales elementos y características de la Tierra y del espacio
	Comprende el funcionamiento de algunas máquinas simples y la relación fuerza- movimiento
Ciencia, Tecnología y Sociedad	Comprende el funcionamiento de diferentes objetos a partir de sus usos y propiedades
	Comprende la diferencia entre varios o diversos tipos de máquinas
	Valora y comprende la necesidad de seguir hábitos para mantener la salud y el entorno

Fuente: Icfes (2014b).

Tabla 11. Competencia: Explicación de Fenómenos - 1° a 5° Grados

Componente	Afirmación
Entorno Vivo	Comprende que los seres vivos dependen del funcionamiento e interacción de sus partes
	Comprende que los seres vivos atraviesan diferentes etapas durante su ciclo de vida
	Comprende que existen relaciones entre los seres vivos y el entorno y que estos dependen de aquellas.
Entorno Físico	Comprende que existe una gran diversidad de materiales que se pueden diferenciar según sus propiedades.
	Comprende que existen diversas fuentes y formas de energía y que esta se transforma continuamente.
	Comprende la estructura básica y el funcionamiento de los circuitos eléctricos
	Comprende y describe la ubicación y las características de la Tierra y algunos cuerpos celestes en nuestro sistema solar
	Comprende el funcionamiento de algunas máquinas simples y la relación fuerza movimiento
Ciencia, Tecnología y Sociedad	Comprende el funcionamiento de diferentes objetos a partir de sus usos y propiedades
	Comprende la diferencia entre varios o diversos tipos de máquinas
	Valora y comprende la necesidad de seguir hábitos para mantener la salud y el entorno
	Comprende la importancia del desarrollo humano y su efecto sobre el entorno

Fuente: Icfes (2014b).

Tabla 12. Competencia: Indagación - 1° a 5° Grados

Componente	Afirmación
Entorno Vivo	Comprende que a partir de la investigación científica se construyen explicaciones sobre el mundo natural
	Utiliza algunas habilidades de pensamiento y de procedimiento para evaluar predicciones
Entorno Físico	Observa y relaciona patrones en los datos para evaluar las predicciones.
	Elabora y propone explicaciones para algunos fenómenos de la naturaleza, basadas en conocimiento científico y de la evidencia de su propia investigación y en la de otros

Fuente: Icfes (2014b).

Prueba de Competencias Ciudadanas. Las Competencias Ciudadanas, se pueden definir como las "capacidades cognitivas, emocionales y comunicativas que, integradas entre sí y relacionadas con conocimientos y actitudes, posibilitan que el ciudadano actúe de manera constructiva en la sociedad" (MEN, 2015). Según los Estándares básicos de Competencias Ciudadanas para la Educación Básica y Media, se clasifican en emocionales, cognitivas, comunicativas e integradoras y se refieren a los ámbitos de convivencia y paz; pluralidad, identidad y valoración de las diferencias, y, participación y responsabilidad democrática (MEN, 2015).

En la prueba, la Vivencia de la Ciudadanía, se entiende como el ejercicio de derechos y deberes y la participación activa en la comunidad, a la cual se pertenece; así que, un ciudadano competente analiza su entorno social y político; tiene presente sus derechos y obligaciones, posee la capacidad

para reflexionar sobre temas sociales, asume su responsabilidad social, se interesa por los asuntos propios de la colectividad, participa en la búsqueda de soluciones a problemas sociales y busca el bienestar de su comunidad (MEN, 2015).

Se evalúa el uso que hace el estudiante de herramientas de pensamiento y su articulación con determinados conocimientos, en torno a cuatro componentes, a saber: conocimiento, argumentación, multiperspectivismo y pensamiento sistémico.

Conocimiento: incluye temas relacionados con los fundamentos de la Constitución de 1991: los derechos y deberes, la estructura del Estado y la participación democrática en el ámbito escolar, municipal, departamental o nacional.

Argumentación: evalúa que el estudiante esté en capacidad para identificar prejuicios, anticipar el efecto de un determinado discurso, comprender las intenciones implícitas en un acto comunicativo, relacionar diversos argumentos, evaluar la validez de generalizaciones y reconocer la confiabilidad de un enunciado y de sus fuentes.

Multiperspectivismo: evalúa la capacidad del estudiante para analizar una problemática desde diversas perspectivas, trascendiendo el punto de vista personal; la prueba requiere, que, a propósito de un conflicto, el estudiante esté en capacidad de comprender su origen, entender qué buscan los diferentes actores, identificar coincidencias y diferencias entre los intereses de los actores, y evaluar la receptividad de las partes ante una posible solución.

Pensamiento Sistémico: analiza la capacidad del estudiante para identificar y relacionar los factores que están presentes en una situación problemática, por lo cual, evalúa la capacidad para identificar las causas, establecer los tipos de factores que están presentes, comprender los tipos

de factores que se enfrentan, comprender los factores que se privilegian en una determinada solución, evaluar la aplicabilidad de una solución y determinar la posibilidad de aplicar una solución en diversos contextos.

Las Tabla 13, contiene las afirmaciones elaboradas para cada componente evaluado en la prueba y para cada competencia.

*Tabla 13. **Competencia: Pensamiento Ciudadano***

Componente	**Afirmación**
Conocimiento	Conoce la Constitución y su función de enmarcar y regular las acciones de las personas y grupos en la sociedad
	Conoce la organización general del Estado y comprende cuáles son las funciones básicas de las ramas del poder público
	Conoce los mecanismos que los estudiantes tienen a su disposición para participar activamente y para garantizar el cumplimiento de sus derechos
Argumentación	Analiza y evalúa la pertinencia y solidez de enunciados o discursos
Multiperspectivismo	Reconoce la existencia de diferentes perspectivas, en situaciones en las que interactúan diferentes partes
	Analiza las diferentes perspectivas presentes en situaciones en las que interactúan diferentes partes
Pensamiento Sistémico	Comprende que los problemas y sus soluciones involucran distintas dimensiones y reconoce relaciones entre estas

Fuente: MEN (2015).

Resumen de las competencias evaluadas en las Pruebas Saber 5°

En la Tabla 14, se resume las competencias que se han evaluado desde 2009 hasta 2017; se puede observar que, en todos los años se evalúa Lenguaje y Matemáticas, y no siempre Ciencias Naturales y Competencias Ciudadanas.

Tabla 14. Competencias Evaluadas y Aspectos Destacados - Grado 5°

Año/Descripción	**Pruebas/Aspectos destacados**
2009	• Prueba de Ciencias Naturales • Prueba de Lenguaje • Prueba de Matemáticas
2012	• Prueba de Ciencias Naturales • Prueba de Lenguaje • Prueba de Matemáticas • Prueba de Competencias Ciudadanas • Estudio de Factores Asociados • A partir de este año, las pruebas se aplican anualmente
2013	• Prueba de Lenguaje • Prueba de Matemáticas • Pruebas de Competencias Ciudadanas
2014	• Prueba de Ciencias Naturales • Prueba de Lenguaje • Prueba de Matemáticas
2015	• Prueba de Lenguaje • Prueba de Matemáticas • Pruebas por Computador • Prueba de Competencias Ciudadanas

2016	Prueba de Ciencias Naturales Prueba de Lenguaje Prueba de Matemáticas
2017	Prueba de Lenguaje Prueba de Matemáticas Esta es la última prueba aplicada

Fuente: elaboración propia con datos Icfes (s.f).

Modelos de Factores Asociados al Desempeño Escolar

En las investigaciones sobre Rendimiento Escolar, se utilizan modelos teóricos para explicar las relaciones existentes entre los elementos que determinan el aprendizaje, que hacen referencia a instituciones educativas, aulas de clase y características de los estudiantes; así pues, "Los estudios de Factores Asociados, surgen de la tradición investigativa sobre eficacia escolar, como una alternativa para buscar explicaciones a la compleja realidad de la Educación, en donde existen diferencias en el aprendizaje entre estudiantes, salones de clases e instituciones educativas" (ICFES, 2016a). Dado el carácter cambiante de la realidad social y educativa, es clave que, los Modelos de Factores Asociados ofrezcan amplitud y flexibilidad, para ajustarse en el tiempo, y que sean claros con respecto al tipo de relaciones, entre desempeño académico y los factores explicativos correspondientes.

En las pruebas Saber 3°, 5° y 9° entre 2012 y 2015, el Icfes recoge información sobre el contexto de los estudiantes, sus familias y la institución educativa en relación con los insumos con los que cuentan las instituciones educativas, los docentes y los estudiantes, y de los procesos de aprendizaje que tienen lugar en la institución educativa y en el aula de clases (Icfes, 2016a).

En general, los Modelos de Factores Asociados, organizan en dimensiones las variables asociadas al aprendizaje cognitivo y reconocen la estructura jerárquica de estas dimensiones. Por su parte, el estudio de

Factores Asociados desarrollado por el Icfes, sigue el modelo Contexto, Insumos, Procesos y Productos - CIPP, debido a sus bondades en términos de completitud, generalidad y simplicidad, en la explicación de las relaciones existentes entre los factores que influencian el aprendizaje (Icfes, 2016a). El modelo contempla las dimensiones universalmente asociadas al aprendizaje, eso sí, sin perder generalidad; permite incluir variables, que atienden a un contexto particular y describe de forma clara y sencilla, la forma en que las variables definidas, influyen en el logro académico de los estudiantes.

En seguida, se describen tres Modelos de Factores Asociados, a saber: Modelo Ecológico, Modelo de Eficacia Escolar de Scheerens y Modelo Contexto, Insumos, Procesos y Productos (CIPP).

Modelo Ecológico

Busca comprender el desarrollo humano en contexto, reconociendo que las diversas etapas del desarrollo están relacionadas y condicionadas por la interacción con el entorno; se representan como un conjunto de círculos con tres niveles de anidamiento: a) contexto, b) institución educativa y c) salón de clases (Icfes, 2016a).

El círculo exterior (Figura 1), representa el medio ambiente regional y local; a él se adscriben las instituciones educativas. Por su parte, las dinámicas culturales, sociales, económicas y políticas del medio ambiente influyen en el sistema educativo y constituyen su entorno. Estos factores inciden en la eficacia de las instituciones educativas y en el proceso de aprendizaje de los estudiantes. A la categoría Contexto, corresponden las condiciones sociales, económicas y demográficas de los estudiantes y sus familias, las cuales inciden en los resultados educativos de los estudiantes (Icfes, 2016a).

Este modelo propone que, los procesos que se desarrollan en el aula están relacionados con elementos sociales del entorno inmediato y del contexto global. En el círculo que representa la institución educativa (Figura 1), se incluyen variables ligadas a los procesos del funcionamiento

organizacional, tales como el liderazgo, el clima laboral y las relaciones humanas, los insumos y servicios con los que cuentan las instituciones y la orientación de los proyectos escolares. Por su parte, en el círculo de la institución educativa, está el aula de clases, en el cual, estudiantes y profesores realizan las actividades encaminadas al aprendizaje y al desarrollo integral de los estudiantes. En su gran mayoría, las variables dan cuenta de los procesos de enseñanza que lideran los docentes; los factores relativos a las relaciones entre estudiantes, docentes y entre ellos; y disponibilidad de materiales pedagógicos para el trabajo en el aula (Icfes, 2016a) (*Figura 1*).

Modelo de Eficacia Escolar de Scheerens

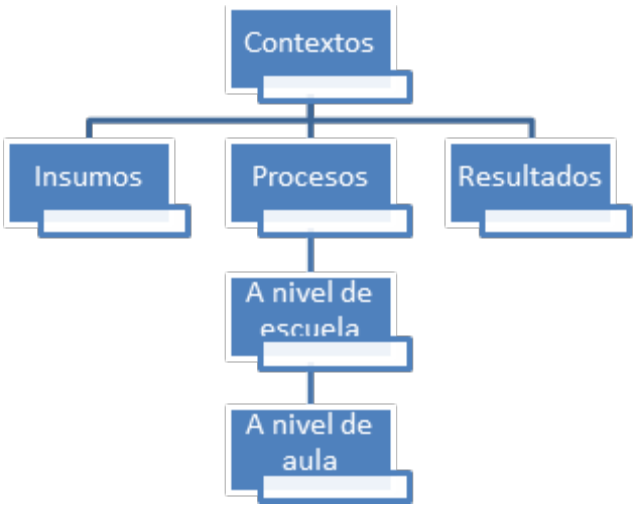
Los Modelos de Eficacia Escolar, comparten una lógica conceptual que relaciona tres aspectos: 1) incluyen la noción ecológica de los procesos educativos; 2) organizan las variables que explican los resultados de aprendizaje en diversos niveles de agregación (contexto, institución educativa y aula); y 3) proponen la direccionalidad de la relación entre las variables explicativas del aprendizaje y el logro del mismo (Icfes, 2016a).

Figura 1. Modelo Ecológico

Fuente: Icfes (2016a)

El Modelo de *Eficacia Escolar* de Scheerens (2000), es reconocido por sintetizar los aspectos claves de los Modelos Conceptuales de Eficacia Escolar; propone las dimensiones: contexto, insumos, procesos y resultados. En *procesos* se encuentra la mayor parte de las variables del modelo, agrupadas en dos categorías: institución educativa y aula de clases (Figura 2).

*Figura 2. **Modelo de Eficacia Escolar de Scheerens***



Adaptado de Icfes (2016a)

La categoría Contextos, incluye los incentivos al aprendizaje que ofrece el Estado, el desarrollo de indicadores macroeconómicos nacionales (consumo, ingreso), el tamaño de la institución educativa, las características de los estudiantes, el tipo de institución y la zona donde se ubica el aula de clases. El Contexto, incide en los procesos que se desarrollan en la institución educativa y en el aula de clases. A nivel de la Institución Educativa, se encuentran los procesos sobre grado de orientación al logro de la política escolar, el liderazgo educativo, el consenso y la planificación docente, la calidad de los contenidos cubiertos y la estructura del proyecto educativo. A nivel de Aula, el modelo incluye, la evaluación de competencias,

las oportunidades de aprendizaje, las expectativas, el monitoreo del progreso de los estudiantes, entre otros. En la categoría Insumos, el modelo considera la experiencia docente, el gasto por estudiante y el apoyo de los padres.

En la categoría Resultados, se ubica el rendimiento académico de los estudiantes, que puede ser ajustada por el rendimiento previo y el nivel socioeconómico del estudiante, que surge de las actividades de enseñanza, dentro de la Institución Educativa.

En general, las variables que explican el aprendizaje de los estudiantes, varían en función del desarrollo económico de los países; de manera que, los factores que promueven el aprendizaje en países con mayores niveles de desarrollo, no necesariamente son los mismos asociados al rendimiento en los países menos desarrollados (Icfes, 2016a). En esta misma perspectiva, el Icfes (2016a) señala que algunos autores manifiestan que, si bien factores de eficacia escolar comparten una base conceptual similar y tienen elementos transversales, en su aplicación, están muy relacionados con la cultura local (Fuller y Clarke, 1994).

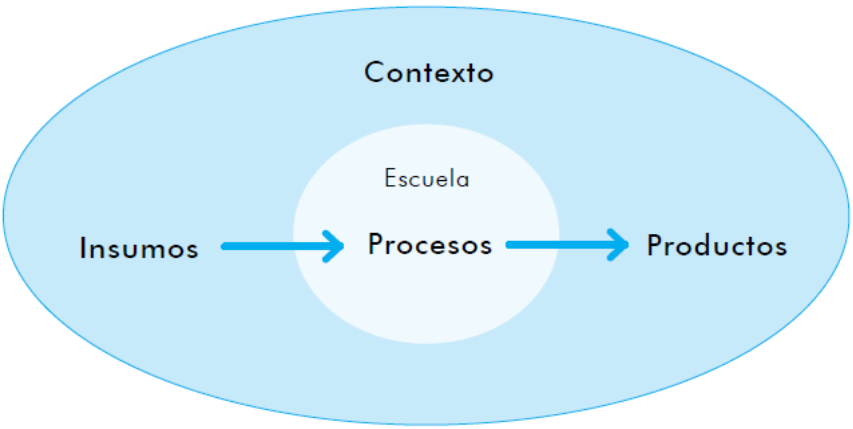
Modelo Contexto, Insumos, Procesos y Productos(CIPP)

El Modelo CIPP, simplifica los modelos de eficacia escolar; se adapta a las condiciones sociales y educativas de un país. Permite seleccionar las variables relevantes para el país, dependiendo de las condiciones socioeconómicas y sus prioridades en materia de política educativa (Icfes, 2016a). (Figura 3).

En la Categoría *Contexto,* se ubican los factores que inciden en el sistema educativo; incluye variables externas a la institución educativa, que afectan los resultados académicos; son relevantes las características sociales, económicas y culturales de los estudiantes y sus familias, que implica un amplio rango de variables, que dependen de la situación social de cada contexto nacional y local.

La Categoría *Insumos*, está enfocada en los recursos de la institución educativa; también se considera el historial educativo de los estudiantes. El modelo propone que, los insumos pueden moldear los procesos. Por su parte, la Categoría *Procesos*, la cual se establece de forma general a nivel de la Institución Educativa, incluye las actividades regulares que se desarrollan en la Institución Educativa y dentro del salón de clases.

Figura 3. Modelo CIPP



Fuente: Icfes (2016a)

La Categoría *Productos*, corresponde a los resultados que se esperan de las actividades que se realizan en la Institución Educativa y en el aula de clases; por su carácter de institución social, uno de los resultados, es el desarrollo cognitivo, social, emocional y ciudadano de los estudiantes. No obstante, a decir del MEN (2006), con las herramientas analíticas y de investigación disponibles, sólo se pueden evaluar aspectos relativos al desarrollo cognitivo, específicamente, relacionadas con el aprendizaje en diversas disciplinas escolares, dejando fuera del alcance del modelo, otras variables que explican el desarrollo del estudiante.

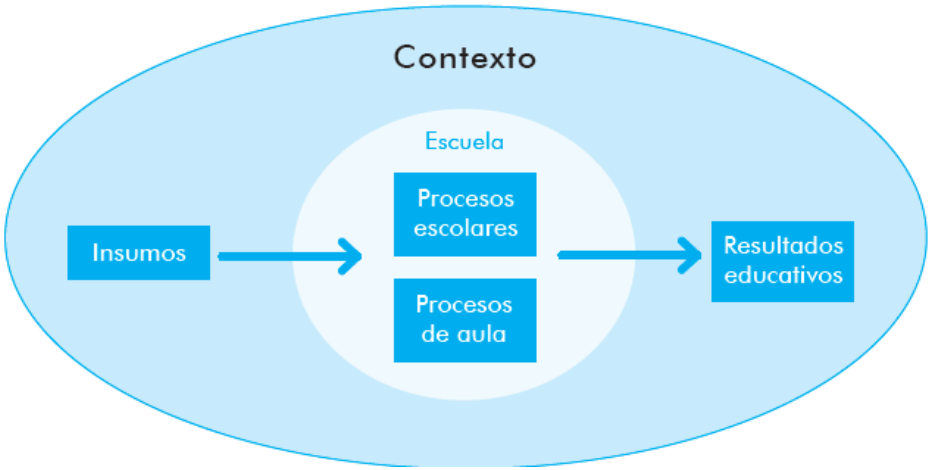
Para efectos del análisis de los factores asociados al desempeño escolar, el Icfes, considerando los criterios de completitud, claridad,

direccionalidad, flexibilidad, escogió el Modelo CIPP, como base para el marco de referencia de los factores asociados al rendimiento académico en las Pruebas Saber 3°, 5° y 9° (Icfes, 2016a).

Factores Asociados a las Pruebas Saber 3º, Saber 5º y Saber 9º

El marco para la prueba Saber 3º, 5º y 9º, sigue el Modelo CIPP, que organiza los factores asociados al aprendizaje en insumos, procesos y resultados, y reconoce que estos elementos están enmarcados en un contexto social, económico y cultural más amplio. Este modelo permite incluir variables generales y específicas, asociadas al sistema educativo local (Figura 4).

Figura 4. Modelo de Factores Asociados adoptado por el Icfes.



Fuente: Icfes (2016a).

A la Categoría Contexto, pertenecen los elementos socioeconómicos y culturales que influyen en la institución, los estudiantes y la comunidad educativa, como también, la ubicación geográfica de la institución educativa,

la modalidad, dependencia administrativa y tamaño del establecimiento. Adicionalmente, pertenecen otros aspectos que tienen relación con características personales de los estudiantes y de sus familias, como el nivel socioeconómico, el género, la motivación y el autoconcepto académico.

En un marco de un estudio de factores asociados, es recomendable hacer seguimiento a las siguientes variables de caracterización individual de los estudiantes: género, edad, trabajo infantil, nivel socioeconómico familiar, violencia en el entorno del hogar, involucramiento parental, pertenencia a un pueblo originario o etnia, necesidades educativas especiales, autoconcepto académico, motivación, estrategias de aprendizaje. Asimismo, en cuanto a la caracterización de las escuelas, se recomienda analizar: zona en la que se ubica la Institución Educativa (urbana o rural), dependencia administrativa, tamaño de la Institución Educativa y el aula de clases, nivel socioeconómico promedio por aula y establecimiento, y violencia en el entorno de la institución educativa.

A la Categoría Insumos, corresponden los recursos de las instituciones, los docentes y los estudiantes. Entre las variables se consideran: a) indicadores de infraestructura escolar (servicios básicos, acceso a computadoras y conexión a internet, entre otras), el tiempo efectivo de aprendizaje y las estrategias de clasificación de estudiantes al interior de las instituciones; b) variables asociadas a los antecedentes escolares, como la asistencia a educación preescolar y la repetición de grado.

A la categoría Procesos de la Institución Educativa y del Salón de Clases, corresponden los procesos de enseñanza y aprendizaje, determinados por las interacciones entre los diversos miembros de la comunidad educativa; también hacen parte de los procesos, las 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.

La categoría Resultados Esperados del Sistema Educativo, tiene que ver con los propósitos de contribuir al desarrollo integral de la población, mediante el proceso de escolarización formal, teniendo en cuenta que la Educación está llamada a apoyar el desarrollo social, emocional, cognitivo y físico de los estudiantes para que estén preparados para la vida en sociedad, sean ciudadanos críticos y responsables, cuenten con habilidades para insertarse en el mercado laboral y encuentren sustento en una actividad productiva.

Factores Asociados al Desempeño Académico en las Pruebas Saber

A partir de los cuestionarios de Factores Asociados aplicados entre 2012 y 2015, es posible obtener información, sobre los aspectos que influyen en el desempeño académico de los estudiantes en la Prueba Saber 3°, 5° y 9°. El estudio de Factores Asociados inició en 2012, con la elaboración de una línea de base; los aplicados en 2012, están dirigidos a rectores, docentes y estudiantes, mientras que los aplicados entre 2013 y 2015, están dirigidos sólo a estudiantes y se denominan, Cuestionarios de Estudiantes.

La Tabla 15, contiene el histórico de los cuestionarios, los cuales recogen información de la dimensión Contexto en cuanto a características de los estudiantes y sus familias.

Tabla 15. Histórico Dimensión Contexto: Características de los Estudiantes y sus Familias

Constructos	Cuestionarios
Género	Cuestionario Sociodemográfico (2009-2015)
Edad	Cuestionario Sociodemográfico (2009-2015)

Trabajo Infantil	Cuestionario Sociodemográfico (2012-2015)
Nivel Socioeconómico Familiar	INSE (2009-2015)
Involucramiento Parental	Cuestionario de Estudiantes (2012)
Violencia en el Entorno del Hogar	Cuestionario de Estudiantes (2012)
Motivación	Cuestionario de Estudiantes (2012)
Estrategias de Aprendizaje	Cuestionario de Estudiantes (2012)
Autoconcepto Académico	Cuestionario de Estudiantes (2012)

Fuente: Elaboración propia con datos de Icfes (2017).

En la Tabla 16, se indica el histórico de recogida de datos, relacionada con las características de las instituciones educativas.

Tabla 16. Histórico Dimensión Contexto: Características de las Instituciones Educativas

Constructos	**Cuestionarios**
Ubicación Geográfica (urbana, rural)	Datos Administrativos
Dependencia Administrativa (oficial, por concesión, privada)	Datos Administrativos
Tamaño del Establecimiento y del Salón de Clases	Datos Administrativos
Nivel Socioeconómico Promedio por Establecimiento	INSE (2009-2015)
Violencia en el Entorno de la Institución Educativa	Cuestionario a Rectores (2012)

Fuente: Elaboración propia con datos de Icfes (2017).

En relación con la Dimensión Insumos, en la Tabla 17, se indica el histórico de recogida de datos correspondiente.

Tabla 17. Histórico Dimensión Insumos

Constructos	Cuestionarios
Antecedentes Escolares: Asistencia a Educación Preescolar y Repetición de Grado	Cuestionario de Estudiantes (2012-2015)
Disponibilidad y Uso de Tecnologías de Información y Comunicación (TIC)	Cuestionario de Estudiantes (2012-2015)
Tiempo Efectivo Dedicado al Aprendizaje	Cuestionario de Estudiantes (2012-2015)
Infraestructura Escolar	Cuestionario de Estudiantes (2012)
Calificación Docente y Conocimiento Profesional	Cuestionario a Docentes (2012)
Materiales Educativos	Cuestionario de Estudiantes (2012)
Programas Escolares Gratuitos	Cuestionario de Estudiantes (2012-2014)

Fuente: Elaboración propia con datos de Icfes (2017).

En cuanto a la Dimensión Procesos, en la Tabla 18, se muestra el histórico de recogida de datos correspondiente.

Tabla 18. Histórico Dimensión Procesos

Constructos	**Cuestionarios**
Estrategias de Clasificación de Estudiantes al Interior de los Establecimientos	Cuestionario a Docentes (2012)
Clima Escolar	Cuestionario de Estudiantes (2012-2015)
Prácticas de Enseñanza Docente	Cuestionario de Estudiantes (2012-2015)
Desarrollo y Colaboración Profesional	Cuestionario a Docentes (2012)
Liderazgo Educativo	Cuestionario a Rectores (2012) Cuestionario a Docentes (2012)

Fuente: Elaboración propia con datos de Icfes (2017).

Finalmente, en la Dimensión Resultados Educativos, en la Tabla 19, se presenta el histórico de recogida de los datos de esta dimensión.

Tabla 19. Histórica Dimensión Resultados Educativos

Constructos	**Cuestionarios**
Aprendizaje en Disciplinas Académicas	Cuadernillo Cognitivo (2012-2015)
Ciudadanía	Cuestionario de Estudiantes (2012-2015)
Bienestar Subjetivo	Cuestionario de Estudiantes (2012-2015)

Fuente: Elaboración propia con datos de Icfes (2017).

Vale decir que, los datos de Factores Asociados y los que corresponden a las Pruebas Cognitivas, se encuentran disponibles para la comunidad académica, en un repositorio organizado y estructurado por el Icfes.

Variables de Contexto Relacionadas con Características de los Estudiantes

En relación con el Contexto, como uno de los Factores Asociados al desempeño académico, los autores Celis, *et al.* (2012) e Icfes (2011), encuentran que, en la Prueba Saber 3°, 5° y 9° de Colombia, el Nivel Socioeconómico tiene un efecto importante. En el mismo sentido, la Unesco (2010, 2015) señala que, en cuanto a las características contextuales, las variables socioeconómicas del estudiantado y de las instituciones educativas, son muy importantes en la explicación de las brechas, sobre desempeño académico en evaluaciones aplicadas a nivel latinoamericano.

Si bien, los establecimientos educativos no tienen control total sobre todas las variables relacionadas con el contexto (Unesco, 2010), mucha literatura sugiere que, algunos aspectos de las prácticas docentes y del clima escolar, influyen de forma significativa en algunas características individuales de los estudiantes, tales como, la motivación para el aprendizaje (Icfes, 2017).

Sin duda, desde la publicación del Informe Coleman (1966), en la explicación de los resultados escolares tienen marcada importancia las características sociodemográficas, económicas y culturales de los estudiantes y sus familias; el efecto positivo de estas variables en el desempeño académico, ha sido ampliamente estudiado y verificado por diversos estudios (OECD, 2011a, 2011b; Unesco, 2015; Unesco, 2010).

Concretamente, las brechas de género en el desempeño académico, en gran medida, se explican por las relaciones de género que prevalecen en

la sociedad y que inciden en los procesos de socialización y educación. En Colombia, los autores Cárcamo y Mola (2012) han analizado, las brechas de género, en el desempeño académico.

La edad, es una variable relacionada con el grado que cursa el estudiante; la inclusión de esta variable permite verificar la correspondencia entre el grado escolar y la edad del estudiante, para analizar la incidencia de casos de extra edad y su relación con otras variables, tales como la repetición de grado escolar (Icfes, 2017).

El trabajo infantil, suele reflejar carencias en el hogar del estudiante y presenta una relación negativa con el rendimiento escolar (Unesco, 2015; Unesco, 2010); pues, los niños que trabajan cuentan con menos tiempo para estudiar; además, su bienestar puede verse deteriorado, dado que no están preparados para asumir tareas laborales. En tal sentido, algunos estudios encuentran que, fuera del horario escolar, el trabajo infantil afecta la capacidad de aprender y estudiar (Gunnarsson, *et al.,* 2006). Dichos autores, utilizan datos de estudiantes de 3° y 4° en nueve países de América Latina; encuentran que, los niños que trabajan, presentan una variación exógena en el costo de oportunidad del tiempo de los niños. Sugieren que el trabajo infantil reduce los puntajes de las pruebas.

En el Índice Socioeconómico de los estudiantes que aplican la Prueba Saber 3°, 5° y 9°, se tiene en cuenta el máximo nivel de escolaridad de los padres, la tenencia de bienes y servicios en el hogar (como agua, electricidad, internet, lavadora, computador, entre otros), indicadores de hacinamiento (como la cantidad de personas que viven con el estudiante y el número de cuartos disponibles) y el material de los pisos y paredes de la vivienda que habita el estudiante (Icfes, 2017).

El entorno económico y cultural de los estudiantes, puede condicionar su desarrollo; pues, la falta de recursos para una alimentación apropiada,

la ausencia de estímulos intelectuales que fortalezcan el desarrollo comprehensivo del lenguaje y el desinterés en la vida escolar, limitan las opciones académicas de los niños y jóvenes (Sirin, 2005).

La participación de los padres, en el proceso formativo de sus hijos, beneficia la formación de habilidades cognitivas y no cognitivas y el desarrollo motivacional; algunos estudios, muestran la estrecha relación entre la participación de los padres y los resultados escolares (Unesco, 2015; Unesco, 2010).

Adicionalmente, los hábitos familiares de lectura, también son importantes, ya que influyen en la formación de capital cultural (Bourdieu y Passeron, 2001) y hacen parte de las prácticas, que inducen a los niños a la lectura y la cultura escolar (Icfes, 2017).

La exposición a altos niveles de violencia y a situaciones de riesgo, como venta y consumo de drogas, tiene implicaciones negativas en el logro académico y en el desarrollo de competencias ciudadanas en los estudiantes (Chaux, 2009).

Tiene implicaciones negativas, en el logro académico y en el desarrollo de competencias ciudadanas de los estudiantes, el hecho, de exponerse a altos niveles de violencia y a situaciones de riesgo, tales como, venta y consumo de drogas (Icfes, 2017). Adicionalmente, varios autores han mostrado que los niños y adolescentes víctimas u observadores de abuso infantil y violencia intrafamiliar, tienen mayor probabilidad de desarrollar problemas de ansiedad, depresión, baja autoestima, irritabilidad y atrasos en el desarrollo del lenguaje, los cuales inciden, en el desempeño escolar y en los procesos socioemocionales de los estudiantes (Icfes, 2017).

Por u parte, la Motivación Intrínseca, corresponde a la motivación producida por gusto o interés en un tema particular y la Motivación

Extrínseca, es producida por reconocimientos, recompensas o características personales (Ryan y Deci, 2000).

En el marco de las Pruebas PISA, la Motivación, se considera parte de los comportamientos o actitudes favorables, frente al aprendizaje que tienen una relación positiva y significativa con los resultados cognitivos de las pruebas aplicadas en 2000 y 2009 (OECD, 2003; OECD, 2010a; OECD, 2016a).

Por su parte, en el estudio de Factores Asociados a las Pruebas Saber 3°, 5° y 9°, se encuentran preguntas para identificar diversos tipos de motivación, y en tal sentido, los estudiantes, reciben cuestionarios diferentes para la evaluación de las pruebas cognitivas; en tal sentido, los estudiantes que responden preguntas cognitivas de las áreas de Ciencias Naturales y Lenguaje, evidencian motivación asociada a estas áreas y los que se evalúan en Matemáticas y Lenguaje, tienen motivación sobre estas áreas.

En cuanto a las estrategias de aprendizaje, las cuales incluyen cualquier comportamiento o práctica que facilite la adquisición, entendimiento y posterior transferencia de nuevos conocimientos o habilidades (Weinstein, *et al.,* 2000), igual que la motivación, hace parte de los comportamientos y actitudes favorables frente al aprendizaje (OECD, 2003; OECD, 2010a; OECD, 2016a). Diversos estudios señalan que existe una relación positiva entre las estrategias de aprendizaje y el desempeño escolar (Guthrie, J. T. y Wigfield, A., 2000; Guthrie, *et al.,*2012). En este contexto, los resultados de PISA 2009, evidencian que, en la mayoría de países participantes, el puntaje de lectura de los estudiantes más desaventajados socio-económicamente, aumentaría si estos tuvieran estrategias efectivas para sintetizar información tal como las tienen estudiantes de mayor nivel socioeconómico (OECD, 2010a). Vale destacar que, las estrategias de trabajo colaborativo, propician habilidades esenciales para la fuerza de trabajo moderna y se han convertido en un componente importante del proceso educativo, mediante el desarrollo de trabajos en grupo (Icfes, 2017).

La confianza del estudiante en el desempeño académico, incide positivamente en la actitud de los estudiantes frente a la institución y puede predecir mayor participación en clase, menores tasas de ausentismo escolar y mayor probabilidad de realizar las tareas asignadas (Green, *et al.,* 2012). En este sentido, la literatura reporta una relación positiva entre diversos indicadores de autoconcepto académico y resultados en pruebas cognitivas (Trautwein y Möller, 2016). Es importante tener en cuenta que, las creencias sobre sí mismos, pueden ser producto de su desempeño académico (Icfes, 2017).

Variables de Contexto Relacionadas con Caracteristicas de las Instituciones Educativas

El nivel socioeconómico de las instituciones educativas y el entorno que los rodea, inciden en la demarcación de brechas sobre el desempeño académico de los estudiantes; en esta perspectiva, el Icfes, dispone de información, sobre algunas características de contexto de las instituciones educativas, así: ubicación geográfica (rural o urbana), tipo de régimen administrativo (oficiales, concesionados, de ampliación de cobertura y privados) y tamaño de las instituciones y del salón de clases en términos del número de estudiantes.

El nivel socioeconómico de la Institución Educativa, es una variable contextual que se reconoce como el principal factor que explica las brechas de aprendizaje entre establecimientos educativos. Vale aclarar que, en cuanto al desempeño académico, el efecto del nivel socioeconómico de la institución es independiente y adicional al efecto que ejerce el nivel socioeconómico del estudiante. A nivel institucional, también tiene efecto sobre el aprendizaje, la presencia de violencia en el contexto del establecimiento educativo (Icfes, 2017). También se indica que, ser testigo de situaciones violentas, se relaciona positivamente con la inasistencia escolar y con el reconocimiento como estudiantes problemáticos (Jenkins y Bell, 1994)

aspectos, que inciden sobre el aprendizaje y tienen efectos negativos sobre los compañeros de clase.

Variables Relacionadas con la Dimensión Insumos

El Icfes, también recoge información sobre los insumos de los establecimientos educativos y de los estudiantes; si bien, estos factores no siempre están relacionados con el aprendizaje, es importante medirlos para analizar las desigualdades en la provisión de recursos entre instituciones educativas y la ubicación geográfica.

Las competencias sobre el uso de tecnologías de información y comunicación (TIC), son cada vez más valoradas y representan un insumo fundamental en los procesos de enseñanza y de aprendizaje y en la gestión de los entornos educativos (Gómez y Macedo, 2010); pues, estas tecnologías presentan diversas alternativas que contribuyen a fortalecer los procesos de enseñanza y de aprendizaje, mediante aplicaciones que posibilitan el trabajo colectivo, los procesos formativos virtuales y la creación de plataformas que facilitan el desarrollo de aplicaciones reales y cotidianas del conocimiento. No obstante, según Icfes (2017), no es clara la relación entre el logro cognitivo y el uso de las TIC, por el acceso a computadores e internet; que más bien, el efecto del acceso y uso de las TIC podría depender del nivel previo de competencias, con el que cuentan estudiantes y profesores; afirmación que tiene implicaciones de política pública y constituye un tema de investigación para la comunidad académica.

Un insumo clave en la formación de competencias cognitivas y no cognitivas, los constituyen los antecedentes escolares de los estudiantes, entre los cuales, son fundamentales las variables relacionadas con competencias adquiridas previamente; por ejemplo, el desempeño académico en etapas previas a la cual se está evaluando, la repetición escolar y la asistencia a educación preescolar (Icfes, 2017).

Sobre la repetición de grado, algunos autores sostienen la premisa de que, la retención escolar, otorga tiempo adicional para que los estudiantes rezagados, logren los aprendizajes esperados y enfrenten menos dificultades en grados superiores (Reschly y Christenson, 2013).

En cambio, otros autores como Holmes, (1989) y Roderick, (1994) no están de acuerdo e indican que, la repetición de grado tiene efectos negativos y acumulativos en el desempeño escolar, que entre menor sea el grado repetido, mayores son las consecuencias negativas en el estudiante; en esta perspectiva, otros estudios muestran que la repetición de grado, afecta negativamente la autoestima, la motivación escolar y la conducta de los estudiantes (Agasisti y Cordero, 2014; Reschly y Christenson, 2013)

La relación positiva entre la asistencia a educación preescolar y el rendimiento académico, es soportada por diversos estudios; pues, la Educación Preescolar, no solo tiene efectos positivos sobre los resultados cognitivos de los estudiantes, sino también sobre la movilidad social y la reducción de la pobreza. Investigaciones asociadas al efecto de esta variable sobre el desempeño escolar, también han analizado el impacto que tiene la inversión en Educación sobre diferentes grupos de estudiantes en cuanto a la asistencia a Educación Preescolar, y han concluido que, es más eficiente cuando existen intervenciones tempranas de calidad y focalizadas en quienes más lo necesitan (Heckman, 2006). Asimismo, en otros proyectos, se ha investigado por la existencia de momentos críticos y sensibles para la formación de habilidades cognitivas y no cognitivas, de lo cual, han concluido que, las habilidades cognitivas son más maleables en edades tempranas (Cunha y Heckman, 2007).

La relación entre el tiempo efectivo de aprendizaje y los resultados académicos, ha sido ampliamente estudiada en diversas investigaciones y concluyen que existe una relación significativa y positiva (Patall *et al*., 2010). Bajo esta perspectiva, el tiempo efectivo dedicado al aprendizaje, es una

variable de difícil medición, por lo que suelen utilizarse aproximaciones que, además del manejo y la organización del tiempo de clase, consideran el tiempo que los estudiantes y docentes no están en el salón de clases, el tiempo no dedicado a la enseñanza por un ambiente de disciplina inapropiado o por factores externos al control del docente, tales como el ruido y las condiciones climáticas (Icfes, 2017). Es importante tener en cuenta que, la información que se utiliza para aproximarse al tiempo efectivo de aprendizaje puede estar relacionada con las Dimensiones Contexto y Procesos; por ejemplo, el ausentismo escolar y la impuntualidad de los estudiantes, puede ocurrir por las características geográficas o por situaciones de orden público en la zona de residencia del estudiante o en el área de ubicación del establecimiento educativo. De igual manera, la pérdida de tiempo, debida a la indisciplina de los estudiantes, puede explicarse en gran medida, por la capacidad del docente para establecer y hacer cumplir normas y acuerdos en el salón de clases.

La calidad docente, es uno de los factores que más influye en el aprendizaje de los estudiantes; por su calidad, algunos profesores pueden lograr que todos los estudiantes aprendan, independientemente de sus características; mientras que otros, por baja calidad, perjudican los aprendizajes estudiantes, especialmente en los primeros años de enseñanza (OECD, 2016a; Unesco, 2015; Unesco, 2010).

Los materiales educativos, refieren la tenencia de libros de texto y materiales de apoyo, la disponibilidad de útiles escolares, espacios adecuados para el aprendizaje, dispositivos tecnológicos con acceso a internet, laboratorios y ayudas audiovisuales (OECD, 2010b). Si bien, la disposición de estos materiales en la Institución Educativa y en la familia, no garantiza resultados educativos sobresalientes, su ausencia sí tiene efecto en el proceso de aprendizaje, debido a que son materiales que facilitan la interacción con los temas de estudio y el desarrollo de trabajos, orientados a la profundización de los logros de aprendizaje (Icfes, 2017).

La disponibilidad y calidad de materiales educativos disponibles en las instituciones educativas de los países desarrollados, es reconocida como un factor con poca influencia en el logro cognitivo de los estudiantes, mientras que, en países subdesarrollados, se reconoce que estos insumos explican un porcentaje considerable de la varianza en el desempeño académico (Icfes, 2017); hallazgos que pueden ocurrir debido a la heterogeneidad en la disponibilidad y calidad de materiales educativos de los países no desarrollados. En este sentido, los resultados de PISA 2009, indican que, quienes leen textos largos y complejos, obtienen mejores resultados en la Prueba de Lectura.

La infraestructura escolar, para algunos autores, se ha reconocido como un constructo de rol secundario en la adquisición de competencias cognitivas, afirmación debatida por diversos autores en el contexto de países subdesarrollados, donde la disponibilidad y calidad de la infraestructura escolar parece tener mayor importancia para el desempeño académico de los estudiantes (Ibidem). Esta variable presenta una relación positiva con el nivel socioeconómico del estudiante y de la institución educativa. Por su parte, Unesco (2015) evidencia una relación positiva entre la infraestructura escolar y los resultados de aprendizaje, siempre que los servicios y espacios físicos se utilicen como apoyo significativo en los procesos de enseñanza.

Los programas escolares gratuitos representan un apoyo fundamental para los estudiantes; sin embargo, los recursos entregados, no siempre están orientados a promover el aprendizaje. Su asignación, suele obedecer a criterios de focalización en instituciones que atienden poblaciones socioeconómicamente desaventajadas, lo que hace necesario aislar el efecto del nivel socioeconómico. Finalmente, la participación en estos programas impacta de forma directa la asistencia escolar, siendo necesario tomar medidas que eviten confundir su influencia en el rendimiento escolar, con el efecto de posibles variaciones en la asistencia escolar. Entre los programas escolares gratuitos, se encuentran los que

reducen los costos de asistencia, los que reducen los costos de útiles y uniformes, y los que introducen beneficios como desayunos o controles médicos gratuitos (Murnane y Ganimian, 2014).

Los programas asociados a alimentación, tienen un efecto positivo en el rendimiento escolar; se destaca que la magnitud y consistencia de este efecto, es mayor en niños que en jóvenes adolescentes y en estudiantes en riesgo de malnutrición.

Variables de Procesos Escolares

Los procesos al interior de la institución educativa y del salón de clases son la clave para el aprendizaje, puesto que es allí, donde se producen las interacciones entre docentes y alumnos que fomentan los procesos cognitivos.

En relación con las políticas de admisión de estudiantes y su ubicación en los salones de clase, obedecen a características del sistema educativo, tales como, el nivel de estratificación escolar; estrategias que están condicionadas por la forma en que se establecen límites socioeconómicos y culturales en el territorio y en las instituciones; de modo que, las instituciones que atienden poblaciones más diversas en términos socioeconómicos y culturales, tienen mayores oportunidades de clasificar estudiantes según diversas características (Icfes, 2017). Según OECD (2016b) Colombia, es uno de los países que presentan mayor estratificación escolar, es decir, la población de estudiantes que asiste a un mismo establecimiento, es muy similar en términos socioeconómicos.

Investigaciones cuantitativas sobre este tema, se han orientado a estudiar la relación entre la agrupación de estudiantes según la habilidad cognitiva y el desempeño académico, frente a lo cual, algunos autores concluyen, que la relación es positiva, debido a la posibilidad de diseñar

planes de aula que se adapten a las necesidades de aprendizaje de los estudiantes, y otros concluyen que la relación es negativa por la pérdida de aprendizaje, causada por la ausencia del efecto de pares en salones con estudiantes homogéneos, en términos de habilidad cognitiva (Betts y Shkolnik, 2000).

Según el Icfes (2017), en estudios realizados en Gran Bretaña, por Kerckhoff, (1986) y Estados Unidos, por Hoffer (1992), se encuentra que, los estudiantes pertenecientes a salones de alto rendimiento escolar, presentan mejor desempeño que los estudiantes promedio de un salón heterogéneo en términos de habilidad; y que, los estudiantes pertenecientes a un salón con bajos niveles de aprendizaje, muestran menor desempeño que los estudiantes promedio de un salón no diferenciado según niveles de habilidad.

El clima escolar, abarca un conjunto de normas y valores que comparten los actores escolares, la calidad de relaciones al interior de la institución educativa y la atmósfera general del establecimiento educativo. Según OECD (2005), el clima escolar, es la variable que explica el mayor porcentaje de las diferencias de desempeño académico entre instituciones educativas. Un clima escolar positivo, está relacionado con el desarrollo favorable de niños y jóvenes, la reducción de riesgos de salud, el logro del aprendizaje y mayores tasas de graduación y retención de profesores. Por su parte, los ambientes positivos de aula, promueven el nivel de involucramiento y motivación de los estudiantes en el aprendizaje y fomentan el desarrollo de prácticas cooperativas, de confianza y respeto (Icfes, 2017). En tal sentido, la Unesco (2015), confirma a nivel latinoamericano, la existencia de una relación positiva entre el clima escolar y el logro académico.

Por su parte, la presencia efectiva de normas y acuerdos en el salón de clases, posibilita la existencia de un ambiente apropiado para los procesos de enseñanza y de aprendizaje. Por el contrario, la ausencia de normas

fomenta la generación de factores de distracción, creando un ambiente inapropiado; en realidad, es fundamental para el aprendizaje y tiene relación con la capacidad de los docentes para administrar el salón de clases (OECD, 2016b). En cuanto a la relación entre el nivel socioeconómico del establecimiento y la presencia efectiva de normas y acuerdos, al parecer, existe evidencia que sustenta la ausencia de estratificación escolar; es decir que, en promedio, los colegios con mayor nivel socioeconómico, no presentan índices de disciplina superiores (menos interrupciones, ruido, mayor nivel de atención) a los observados en colegios con bajo nivel socioeconómico (Arum y Velez, 2012).

En torno a la Dimensión Procesos Escolares, otras variables que también inciden en el desempeño académico son las siguientes (Icfes, 2017):

Relaciones interpersonales, las cuales centran la atención en la calidad y el tipo de relaciones entre estudiantes y docentes, por ser las que inciden de forma directa en los procesos de enseñanza y aprendizaje.

a. Percepción frente al establecimiento educativo, en la cual, se considera las creencias, percepciones y actitudes de los estudiantes frente a la institución educativa.

b. Prácticas de enseñanza docente: activación cognitiva, seguimiento al aprendizaje y estructura y organización de clase; pues, los enfoques clásicos de la investigación instruccional, conciben el aprendizaje como un proceso individual, producto de la instrucción directa y de la transmisión de conocimiento del docente al estudiante; en cambio, los enfoques constructivistas, conciben que el aprendizaje es un proceso donde el estudiante crea de forma activa, conocimiento y competencias, en un contexto social y cultural.

c. Organización y estructura de clase, se refiere a la capacidad del profesor para introducir y mantener una clase estructurada y bien orientada.

d. Desarrollo y colaboración profesional, se refiere a la realización de actividades conjuntas que fomenten la creación de herramientas y recursos para proveer enseñanza de calidad cuyo objetivo, es mejorar las prácticas pedagógicas y asegurar que los estudiantes desarrollen las habilidades y conocimientos establecidos en el currículo.

e. Liderazgo educativo, que hace referencia a la capacidad para guiar y adelantar programas de desarrollo profesional, definir metas escolares de aprendizaje, monitorear y garantizar que las lecciones de clase sean acordes a las metas de aprendizaje, sugerir cambios orientados a mejorar las prácticas docentes y ayudar a resolver problemas que puedan surgir entre los integrantes de la comunidad escolar (Icfes, 2017).

Variables Relacionadas con Resultados Educativos

Es natural vincular como producto del proceso educativo a los resultados de pruebas estandarizadas, los cuales miden el aprendizaje cognitivo en diversas áreas, pero en realidad, el propósito de la educación, no es sólo el desarrollo cognitivo, sino la formación integral de los estudiantes, que tiene relación con el desarrollo de actitudes y destrezas, que posibiliten prácticas de convivencia y ciudadanía adecuadas, además, que contribuyan a que las personas ingresen al mercado laboral y aporten al desarrollo de la sociedad (OECD, 2016a).

En este sentido, la Prueba Saber 3°, 5° y 9°, genera resultados cognitivos en las áreas de Lenguaje, Matemáticas, Ciencias Naturales y Pensamiento Ciudadano; y no cognitivos, en relación con convivencia y paz, participación y responsabilidad democrática, y pluralidad y valoración de las diferencias. Así pues, existe un esfuerzo importante por incluir en la medición de resultados de las Pruebas Saber, las habilidades ciudadanas (Icfes, 2017).

Finalmente, se puede afirmar que, el objetivo la Prueba Saber 5°, sobrepasa la obtención de resultados cognitivos, dado que recoge información sobre características y comportamiento de factores asociados al aprendizaje; por lo cual, se constituyen en un recurso para la toma de decisiones de política educativa a nivel nacional, regional y al interior de las instituciones educativas; pues, los cuestionarios de factores asociados aplicados entre 2012 y 2015, recogen información sobre el contexto familiar de los estudiantes y de la institución, sobre los insumos con los que cuentan los establecimientos educativos, los docentes y los estudiantes y los procesos de aprendizaje, que tienen lugar en el establecimiento educativo y en el salón de clases.

Estudios Generales

En el estudio Geometría y Medición en las Pruebas Saber 5°: ¿qué evalúan? Los autores Marmolejo, *et al.* (2019) estudian la alineación entre el contenido geométrico y métrico que se evalúa en las Pruebas Saber 5° de Colombia, y la alineación de las mismas con las sugerencias de los Lineamientos y Estándares curriculares del País. Trabajaron con datos cualitativos y los analizaron con base en las categorías pensamientos y conocimientos en cuanto al estudio de la geometría y la medición.

El objetivo de la investigación fue determinar el nivel de alineación entre el contenido evaluado en las preguntas de las Pruebas Saber y el expuesto en los documentos oficiales colombianos. La atención se centró en las preguntas de las Pruebas Saber 5°, las cuales evalúan objetos, propiedades, relaciones y pensamientos geométricos y métricos; asimismo, en los referentes conceptuales que determinan la enseñanza de las Matemáticas en Colombia (Marmolejo *et al.,* 2019).

Las Pruebas Saber 5º, se aplican periódicamente a estudiantes de Educación Básica de todo el país. Son de carácter obligatorio y censal,

según lo establece la Ley 715 del 2001. Buscan caracterizar el desarrollo de las competencias básicas, tanto en Lenguaje como en Matemáticas y Ciencias. A partir del 2002, se evalúa a los estudiantes de Grado Quinto de los calendarios A y B de todas las instituciones educativas colombianas y, desde el 2012, su aplicación es anual.

Para el análisis, se utilizaron las categorías pensamientos y conocimientos matemáticos, asociados al estudio de la geometría y la medición. La primera, asume el Pensamiento Espacial como "el conjunto de los procesos cognitivos, mediante los cuales se construyen y se manipulan las representaciones mentales de los objetos del espacio" (MEN, 1998). El métrico, por su parte, como "la comprensión general que tiene una persona sobre las magnitudes y las cantidades, su medición y el uso flexible de los sistemas métricos o de medidas en diferentes situaciones" (MEN, 2006).

Los autores concluyen, que "la Prueba tiende a evaluar desarticuladamente, los pensamientos espacial y métrico, ignorar su vínculo con otros pensamientos y privilegiar, en algunas preguntas, competencias desarticuladas de un contenido específico" (Marmolejo, *et al.,* 2019). Asimismo, encontraron inadecuados niveles de alineación con las sugerencias expresadas en los Lineamientos y Estándares nacionales (Marmolejo *et al.,* 2019).

CAPÍTULO II
EFECTOS EN EL DESEMPEÑO ACADÉMICO DE LAS PRUEBAS SABER 5° EN EL DEPARTAMENTO DE NARIÑO

Evaluar el desempeño académico de los estudiantes que pertenecen a las instituciones de Educación Primaria, es un proceso importante, debido a que permite obtener evidencias sólidas de la calidad de la Educación que ofrecen dichas instituciones en cada departamento, municipio y región del país, e identificar los factores que afectan el desempeño de sus estudiantes, que constituye un insumo de alta consideración para el diseño de políticas públicas enfocadas al mejoramiento de los procesos educativos.

En este capítulo, inicialmente, se describen tanto las características socioeconómicas de los estudiantes del departamento de Nariño como las institucionales de los colegios a los cuales pertenecían los mismos, las cuales pueden generar posibles brechas en el nivel de desempeño de los estudiantes en las pruebas Saber 5º.

Posteriormente, se establece un ranking según el municipio y la subregión donde están ubicadas las instituciones de Educación Primaria del departamento de Nariño, cuyos estudiantes presentaron las pruebas Saber 5º, evaluadas en 2017. Igualmente, se determina el efecto que tiene en el desempeño de dichas competencias, cada una de las variables socioeconómicas: género, edad, nivel educativo del padre, nivel educativo de la madre, hacinamiento, condiciones de TIC y electrodomésticos, haber reprobado el Grado 5, y de las variables institucionales: tipo de colegio, naturaleza, carácter, área de ubicación y jornada.

Para obtener los resultados anteriores, se utilizó el Estadístico d de Cohen (1998), el cual, mediante diferencias estandarizadas, entre las diferentes categorías de las variables observadas, de las medias

de los puntajes obtenidos en cada prueba, permite establecer lo que denominaremos el tamaño del efecto de cada variable sobre el rendimiento en las pruebas Saber 5°. Se realiza una clasificación, a partir del estadístico anterior, tomando como referente el grupo o categoría en cada variable, con más alto promedio alcanzado de cada competencia.

La expresión que permite calcular el Estadístico d de Cohen, a partir de las diferencias estandarizadas entre los promedios en cada una de las categorías de cada variable evaluada frente a la categoría o grupo que alcanzó el máximo puntaje promedio en cada una de las competencias evaluadas, y que se toma como referencia, es la siguiente:

$$d_i = \frac{\bar{X}_0 - \bar{X}_i}{S_{i0}} \quad (1)$$

donde: S_{i0} $\sqrt{\frac{(n_0-1){S_0}^2+(n_i-1){S_i}^2}{n_0+n_i-2}}$.

d_i = Tamaño del efecto de la *categoría i* de la variable, con relación a la categoría de referencia.

$\bar{X}_0$ = Media de la categoría de referencia (de mayor puntaje).

$\bar{X}_i$ = Media en la categoría *i* de la variable.

${S_0}^2$ = Varianza de la categoría de referencia.

${S_i}^2$ = Varianza de la categoría *i* de la variable.

N_0 = Número de estudiantes que presentaron las pruebas en la categoría de referencia.

N_i = Número de estudiantes que presentaron las pruebas en categoría i de la variable.

Para la clasificación correspondiente en cada variable, se tiene en cuenta la siguiente escala que propone Cohen (1998), para interpretar las diferencias en unidades estándar, obtenidas con el estadístico de la expresión (1): [0.0,0.2] trivial o irrelevante (grupo A); (0.2,0.5] pequeña (grupo B), (0.5,0.8] moderada (grupo C); y (0.8,∞) grande (grupo D).

Características Socioeconómicas

En el departamento de Nariño, 24.062 estudiantes de Grado 5° presentaron las pruebas Saber 5° durante el año 2017. La Tabla 20, presenta las características socioeconómicas de dichos estudiantes.

*Tabla 20. **Características Socioeconómicas de los Estudiantes Nariñenses que Presentaron las Pruebas Saber** 5°*

Variable socioeconómica		**N**	**%**
Genero	Masculino	12.739	52,9%
	Femenino	11.323	47,1%
Edad	9 años o menos	924	3,8%
	10 años	9.690	40,3%
	11 años	7.660	31,8%
	12 años o mas	5.788	24,1%
Educación del Padre	Primaria	9.610	39,9%
	Bachillerato	10.070	41,9%
	Técnico o Tecnólogo	1.105	4,6%
	Universitario – Pregrado	1.457	6,1%
	Universitario – Posgrado	1.820	7,6%
Educación de la Madre	Primaria	8.937	37,1%
	Bachillerato	10.100	42,0%
	Técnico o Tecnólogo	1.289	5,4%
	Universitario – Pregrado	1.724	7,2%
	Universitario – Posgrado	2.012	8,4%

Hacinamiento	Hacinamiento crítico	231	1,0%
	Hacinamiento medio	2.554	10,6%
	Sin hacinamiento	21.277	88,4%
TIC	Malo	13.820	57,4%
	Regular	3.168	13,2%
	Bueno	7.074	29,4%
Electrodomésticos	Malo	11.858	49,3%
	Regular	4.675	19,4%
	Bueno	7.529	31,3%
Reprobado	No	17.174	71,4%
	Si	6.888	28,6%
Total		24.062	100,0%

Fuente: elaboración propia.

Según la tabla anterior, el 52.9% de los estudiantes que presentaron la prueba, son de Sexo masculino, mientras que el 47,1% son mujeres; el 72,1% de los participantes están entre los 10 y 11 años. Los Niveles Educativos de las madres y los padres de los estudiantes, son un tanto similares, aunque es destacable que las madres presentan niveles un tanto superiores que los padres, el 21,0% de las primeras tiene estudios superiores (tecnólogos, pregrado, posgrado), mientras que en los segundos este porcentaje es de 18,3%. El 11,6% de los estudiantes presenta algún nivel de Hacinamiento en sus viviendas; solamente el 29,4% manifiesta tener acceso a unas TIC buenas condiciones, y de igual manera el únicamente 31,3% cuenta con electrodomésticos en buen estado. El 28,6% de los estudiantes que presentaron las pruebas, ha reprobada el Grado 5.

Características Institucionales

Como se observa en la Tabla 21, el 96,1% de los estudiantes que presentaron las pruebas Saber 5° durante el año 2017, pertenecían a colegios de tipo mixto y menos del 4% a colegios solamente de mujeres o solamente de hombres; el 93,1% pertenecían a colegios oficiales, el 66,1% realizaron sus estudios en colegios de carácter académico, el 9,4% en Técnico y el 24,4% restante en colegios de tipo Técnico/Académico; el 42.7% de las instituciones estaban ubicadas en el área rural y 57,3% en la urbana, en su gran mayoría operando en jornada de la mañana (80,4%), el 13,4% en la tarde y el 6,2% restante en otras jornadas. Aparte de Pasto cuya participación de estudiantes fue del 25,2%, los municipios de mayor participación en las Pruebas Saber 5° de 2017 fueron San Andrés de Tumaco (16.0%) e Ipiales (9.5%). Finalmente, la mayoría de los estudiantes que presentaron las pruebas, pertenecían a las subregiones Pasto (25,2%), Obando (18,5%) y Pacífico Sur (16,7%).

*Tabla 21. **Características Institucionales de los Estudiantes Nariñenses que Presentaron las Pruebas Saber** 5°*

Variables institucionales		**N**	**%**
Tipo Colegio	Femenino	942	3,9%
	Masculino	3	0,0%
	Mixto	23.117	96,1%
Naturaleza	No Oficial	1.672	6,9%
	Oficial	22.390	93,1%
Carácter	Académico	15.910	66,1%
	Técnico	2.271	9,4%
	Técnico/Académico	5.881	24,4%
Área Ubicación	Rural	10.269	42,7%
	Urbano	13.793	57,3%

Jornada	Completa	177	0,7%
	Mañana	19.338	80,4%
	Sabatina	8	0,0%
	Tarde	3.226	13,4%
	Única	1.313	5,5%
Municipio	Albán	133	0,6%
	Aldana	150	0,6%
	Ancuya	77	0,3%
	Arboleda	122	0,5%
	Barbacoas	719	3,0%
	Belén	95	0,4%
	Buesaco	354	1,5%
	Chachagui	201	0,8%
	Colón	143	0,6%
	Consacá	124	0,5%
	Contadero	67	0,3%
	Córdoba	226	0,9%
	Cuaspud	130	0,5%
	Cumbal	508	2,1%
	Cumbitara	122	0,5%
	El Charco	414	1,7%
	El Peñol	89	0,4%
	El Rosario	102	0,4%
	El Tablón de Gómez	251	1,0%
	El Tambo	219	0,9%
	Francisco Pizarro	160	0,7%
	Fúnes	90	0,4%
	Guachucal	250	1,0%

Municipio	Guatarilla	138	0,6%
	Gualmatán	110	0,5%
	Iles	115	0,5%
	Imués	88	0,4%
	Ipiales	2283	9,5%
	La Cruz	269	1,1%
	La Florida	122	0,5%
	La Llanada	68	0,3%
	La Tola	135	0,6%
	La Unión	455	1,9%
	Leiva	159	0,7%
	Linares	153	0,6%
	Los Andes	82	0,3%
	Magui	186	0,8%
	Mallama	127	0,5%
	Mosquera	145	0,6%
	Nariño	52	0,2%
	Olaya Herrera	401	1,7%
	Ospina	97	0,4%
	Pasto	6074	25,2%
	Policarpa	196	0,8%
	Potosí	167	0,7%
	Providencia	72	0,3%
	Puerres	142	0,6%
	Pupiales	218	0,9%
	Ricaurte	230	1,0%
	Roberto payán	170	0,7%
	Samaniego	392	1,6%
	San Andrés de Tumaco	3860	16,0%

Municipio	San Bernardo	105	0,4%
	San Lorenzo	317	1,3%
	San pablo	207	0,9%
	San Pedro de Cartago	94	0,4%
	Sandoná	323	1,3%
	Santa Bárbara	153	0,6%
	Santacruz	119	0,5%
	Sapuyes	95	0,4%
	Taminango	245	1,0%
	Tangua	122	0,5%
	Túquerres	663	2,8%
	Yacuanquer	167	0,7%
Región	Centro	664	2,8%
	Guambuyaco	458	1,9%
	Juanambú	1.342	5,6%
	La Sabana	1.081	4,5%
	La Cordillera	824	3,4%
	Los Abades	583	2,4%
	Obando	4.456	18,5%
	Occidente	677	2,8%
	Pacífico Sur	4.020	16,7%
	Pasto	6.074	25,2%
	Piedemonte Costero	357	1,5%
	Rio Mayo	1.203	5,0%
	Sanquianga	1.248	5,2%
	Telembí	1.075	4,5%
Total		24.062	100,0%

Fuente: elaboración propia.

Resultados de los Estudiantes Nariñenses en las Pruebas de Lenguaje y Matemáticas del Saber 5º

Figura 5. Resultados del nivel de desempeño de los estudiantes nariñenses en la competencia de Lenguaje de las Pruebas Saber 5°

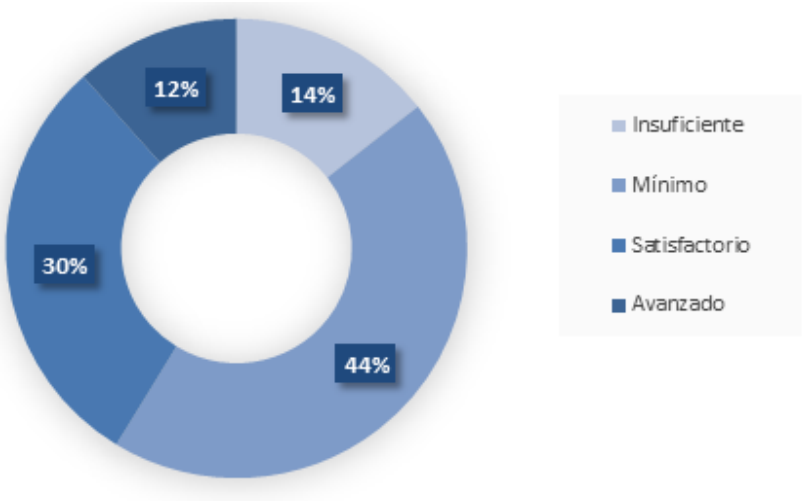
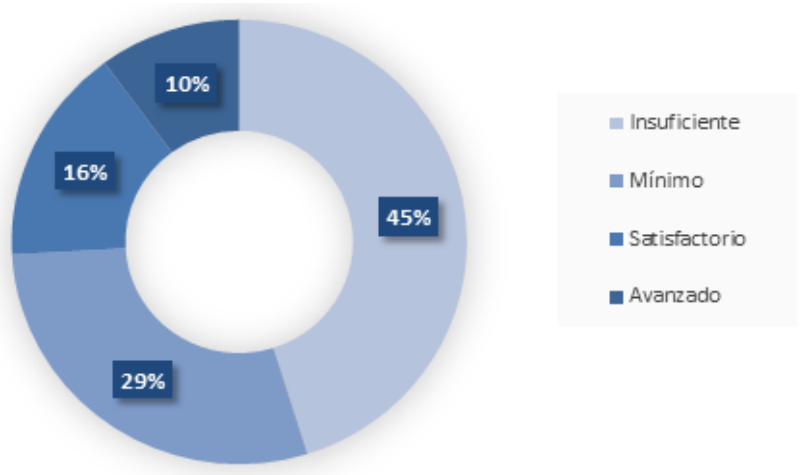


Figura 6. Resultados del nivel de desempeño de los estudiantes nariñenses en las competencia de Matemáticas de las Pruebas Saber 5°



Como se observa en la *Figura 5* y *Figura 6*, en la Prueba de Lenguaje, el 41,3% de los estudiantes alcanzó un rendimiento al menos satisfactorio, mientras que el 58,7% tiene un rendimiento insuficiente o mínimo. El desempeño de los estudiantes del departamento de Nariño en la Prueba de Matemáticas es muy inferior a la del Lenguaje, solamente el 25,9% tuvo un rendimiento cuando menos satisfactorio, mientras que el 74,1% obtuvo un rendimiento insuficiente o mínimo.

Efecto Localidad en las Pruebas de Lenguaje y Matemáticas en el departamento de Nariño

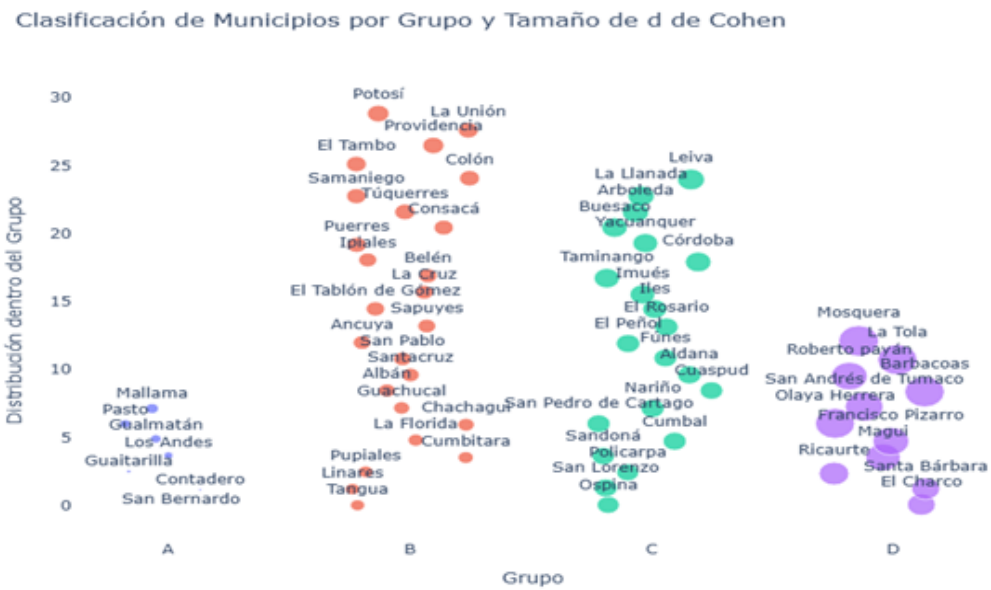
Para evaluar el efecto de la localidad en el rendimiento en las competencias de Lenguaje y Matemáticas de los estudiantes del departamento de Nariño, se elaboró un ranking según el municipio y la subregión, donde se ubican las diferentes Instituciones de Educación Primaria del departamento, que participaron en las pruebas Saber 5° en el año 2017.

Efecto Localidad en Lenguaje en el Departamento de Nariño

La Figura 7 presenta la clasificación por municipio de los resultados de la competencia de Lenguaje en las Pruebas Saber 5° del año 2017 en el departamento de Nariño. Los estudiantes del Municipio de San Bernardo, obtuvieron el promedio de puntaje más alto en esta competencia, seguido por los puntajes en el Contadero, Guaitarilla, Los Andes, Gualmatán, Pasto y Mallama, con diferencias irrelevantes, según el d de Cohen, lo que los ubica en el grupo A de la competencia observada (posiciones 1 a 7). Los municipios en la Clase B, encabezados por el municipio de Tangua, ocuparon las posiciones 8 a 32 con diferencias pequeñas con respecto al máximo puntaje; en el grupo C están los municipios que ocuparon las posiciones

33 a 53, encabezadas por el municipio de Ospina, presentaron diferencias moderadas y los municipios ubicados en el grupo D, encabezados por el municipio de El Charco, alcanzaron diferencias grandes, ocupando las últimas posiciones de la 54 a 64.

Figura 7. Clasificación de los municipios de Nariño de acuerdo con los resultados en Lenguaje de las Pruebas Saber 5°



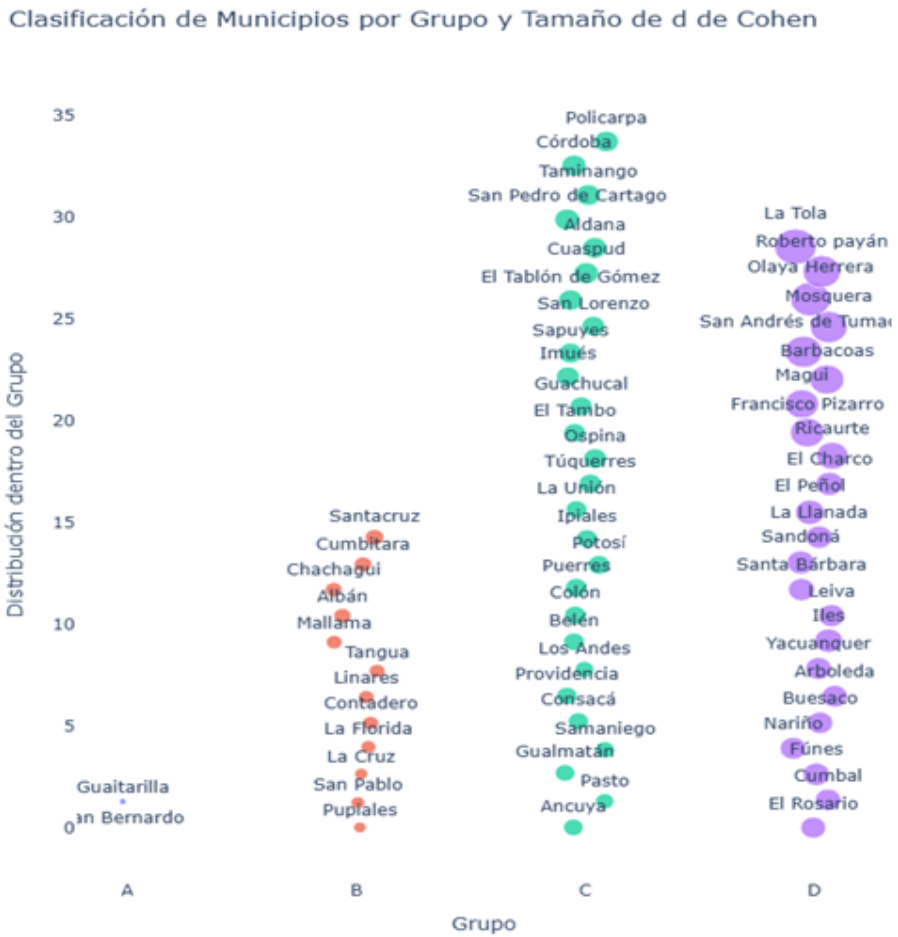
Según la *Figura 8*, en la Prueba de Lenguaje, las subregiones Pasto y Rio Mayo, ocuparon las primeras posiciones ubicándose en el grupo A, mientras que Centro, La Sabana, Los Abades, Obando, Occidente, Guambuyaco, Juanambú, La Cordillera y Piedemonte se ubicaron en el grupo B; Sanquianga, Telembí y Pacífico Sur, ocuparon los últimos lugares en el grupo D, este último con diferencias grandes según el estadístico d de Cohen.

Figura 8. Clasificación de las subregiones de Nariño de acuerdo con los resultados en Lenguaje de las Pruebas Saber 5°

Efecto Localidad en Matemáticas en el Departamento de Nariño

Como se observa en la Figura 9, en el grupo A, se ubicaron en ese orden, los Municipios San Bernardo y Guaitarilla, que también pertenecían al mismo grupo en la Prueba de Lenguaje, ocupando los lugares, primero y tercero en dicha prueba, respetivamente. En el grupo B se ubicaron los municipios que ocuparon los puestos 3 a 14, encabezados por el municipio de Pupiales; en las posiciones 15 a 41 están los municipios que se ubicaron en el grupo C, encabezados por el municipio de Ancuya, con diferencias moderadas respecto a los que se clasificaron en el grupo A; finalmente, ocupando las últimas casillas de la 42 a la 64, con diferencias grandes, están los municipios del grupo D, con El Rosario a la cabeza.

Figura 9. Clasificación de los municipios de Nariño de acuerdo con los resultados en matemáticas de las Pruebas Saber 5°



En la clasificación por subregión del desempeño de la Prueba de Matemáticas, vuelven a ubicarse en el grupo A, Rio Mayo y Pasto, como en la Prueba de Lenguaje, como se observa en la Figura 10 y también Los Abades y Centro, se ubican en el mismo grupo. Las subregiones La Sabana, Obando, Occidente, Juanambú, La Cordillera y Guambuyaco, pertenecen al grupo B; Piedemonte se ubica en el grupo C y ocupando los últimos

puestos, en el grupo D, están Sanquianga, Telembí y Pacífico Sur, igual que en la Prueba de Lenguaje.

Figura 10. Clasificación de las subregiones de Nariño de acuerdo con los resultados en Matemáticas de las Pruebas Saber 5°

Efecto de las Variables Socioeconómicas en el Desempeño de las Competencias de Lenguaje y Matemáticas en las Pruebas Saber 5º del Departamento de Nariño

Se describe a continuación, el efecto de los diferentes factores socioeconómicos, sobre los puntajes obtenidos en las competencias de Lenguaje y Matemáticas de las Pruebas Saber 5° que se presentaron en 2017, en el departamento de Nariño.

Efecto de la prueba de Lenguaje

Como se observa en la Tabla 22, las mujeres y los hombres presentan diferencias no relevantes ($d \leq 0{,}2$) en el desempeño de la Prueba de Lenguaje. Los estudiantes entre los 10 y 11 años tienen un mejor rendimiento, alcanzando diferencias pequeñas frente a los de 9 años o menos y diferencias moderadas frente a los de 12 años o más. Al igual de lo que ocurre a nivel nacional, se observa que los Niveles Educativos de la madre y del padre, son factores que aportan, un efecto importante en los resultados de la Prueba de Lenguaje; estudiantes con padres de niveles de educación universitaria (pregrado o posgrado), presentan un mejor rendimiento en esta competencia, con diferencias estandarizadas pequeñas, frente a los que tienen nivel técnico, bachillerato o primaria.

Las Condiciones de Vida de los estudiantes en el departamento de Nariño, tienen un efecto importante sobre el desempeño en la Prueba de Lenguaje, mejores condiciones de vida redundan en mejores resultados. Estudiantes sin Hacinamiento o con hacinamiento medio en sus viviendas, obtienen mayor rendimiento en Lenguaje, que aquellos que presentan Hacinamiento crítico, aunque con diferencias pequeñas. Se observó, que condiciones buenas en TIC implica mejor desempeño en esta prueba, con diferencias pequeñas; mientras que las condiciones de electrodomésticos no muestran un efecto importante sobre el desempeño. Los estudiantes que reprobaron el Grado 5, mostraron más bajo rendimiento que los que lo aprobaron, con diferencias pequeñas en los promedios estandarizados de la Prueba de Lenguaje.

*Tabla 22. **Efecto de las Variables Socioeconómicas en el Desempeño de Lenguaje en las Pruebas Saber** 5° del Departamento de Nariño*

Variable Socioeconómica		N	Media	DT	d Cohen
Género	Femenino	11.323	312,8	63,3	-
	Masculino	12.739	302,8	61,7	0,16
Edad	9 años o menos	924	296,1	59,5	0,37
	10 años	9.690	318,9	62,1	-
	11 años	7.660	311,2	63,0	0,12
	12 años o mas	5.788	285,5	57,5	0,55
Educación del Padre	Primaria	9.610	303,1	59,2	0,47
	Bachillerato	10.070	304,2	61,4	0,44
	Técnico o Tecnólogo	1.105	311,3	69,5	0,30
	Universitario - Pregrado	1.457	326,3	69,9	0,08
	Universitario - Posgrado	1.820	331,9	68,9	-
Educación de la Madre	Primaria	8.937	303,0	58,6	0,47
	Bachillerato	10.100	302,8	60,7	0,46
	Técnico o Tecnólogo	1.289	314,2	69,8	0,25
	Universitario - Pregrado	1.724	325,5	71,5	0,09
	Universitario - Posgrado	2.012	331,5	68,0	-

Hacinamiento	Hacinamiento crítico	231	279,9	58,5	0,45
	Hacinamiento medio	2.554	304,8	60,0	0,05
	Sin hacinamiento	21.277	308,1	63,0	-
TIC	Malo	13.820	303,4	59,8	0,24
	Regular	3.168	301,3	61,8	0,26
	Bueno	7.074	318,4	67,1	-
Electrodomésticos	Malo	11.858	305,2	58,9	0,15
	Regular	4.675	302,6	62,7	0,18
	Bueno	7.529	314,3	67,6	-
Reprobado	Si	6.888	288,5	56,9	0,43
	No	17.174	315,1	63,3	-

Fuente: elaboración propia.

Efecto en la prueba de Matemáticas

El desempeño de los estudiantes del departamento de Nariño en la Prueba de Matemáticas de las Pruebas Saber 5° en el año 2017, es totalmente similar en los hombres y las mujeres, como se aprecia en la Tabla 23. Contrariamente, los estudiantes entre los 10 y 11 años tienen un mejor desempeño frente a los de 9 años o menos y los de 12 años o más, aunque con diferencias pequeñas en el estadístico de Cohen. Similarmente, el desempeño de los estudiantes en esta competencia es más alto en estudiantes con padres y madres que tienen educación universitaria de pregrado o posgrado, que aquellos estudiantes con padres y madres que tienen estudios técnicos/tecnológicos, bachillerato o primaria, con diferencias estandarizadas pequeñas.

*Tabla 23. **Efecto de las Variables Socioeconómicas en el Desempeño de Matemáticas en las Pruebas Saber** 5° del Departamento de Nariño*

Variable Socioeconómica		**N**	**Media**	**DT**	**d Cohen**
Género	Femenino	11.323	294,3	65,6	0,00
	Masculino	12.739	294,2	67,9	-
Edad	9 años o menos	924	281,7	66,5	0,35
	10 años	9.690	305,5	67,4	-
	11 años	7.660	297,9	65,4	0,11
	12 años o mas	5.788	272,8	62,4	0,50
Educación del Padre	Primaria	9.610	292,6	65,1	0,33
	Bachillerato	10.070	290,1	65,7	0,37
	Técnico o Tecnólogo	1.105	297,7	71,9	0,23
	Universitario - Pregrado	1.457	306,7	71,7	0,11
	Universitario - Posgrado	1.820	314,4	71,3	-
Educación de la Madre	Primaria	8.937	292,2	64,9	0,33
	Bachillerato	10.100	289,5	65,4	0,37
	Técnico o Tecnólogo	1.289	299,2	71,6	0,21
	Universitario - Pregrado	1.724	306,7	71,7	0,10
	Universitario - Posgrado	2.012	313,7	70,2	-

Hacinamiento	Hacinamiento crítico	231	270,9	57,6	0,36
	Hacinamiento medio	2.554	291,7	64,3	0,05
	Sin hacinamiento	21.277	294,8	67,2	-
TIC	Malo	13.820	292,4	64,9	0,13
	Regular	3.168	287,5	65,0	0,20
	Bueno	7.074	301,1	70,8	-
Electrodomésticos	Malo	11.858	294,2	64,6	0,06
	Regular	4.675	288,1	65,5	0,15
	Bueno	7.529	298,2	70,7	-
Reprobado	Si	6.888	277,1	61,5	0,37
	No	17.174	301,2	67,7	-

Fuente: elaboración propia.

Respecto a las Condiciones de Vida de los estudiantes que presentaron la prueba, se observó que más bajo desempeño en la competencia de Matemáticas en aquellos estudiantes que presentan Hacinamiento crítico en sus viviendas presentan que los que viven con hacinamiento medio o sin hacinamiento. Las condiciones de las TIC y los Electrodomésticos no generan diferencias estandarizadas relevantes en los puntajes promedios obtenidos en Matemáticas. Los estudiantes que reprobaron el Grado 5, obtuvieron desempeños más bajos que los que lo aprobaron.

Efecto de las Variables Institucionales en el Desempeño de las Competencias de Lenguaje y Matemáticas en las Pruebas Saber 5º del Departamento de Nariño

Se presenta a continuación, el efecto de las variables institucionales, en el desempeño de las competencias de Lenguaje y Matemáticas de las Pruebas Saber 5° del año 2017, de los estudiantes del departamento de Nariño.

Efecto en la prueba de Lenguaje

Se observa en la Tabla 24, que el tipo de colegio, femenino, masculino o mixto, no genera diferencias importantes, en el rendimiento de los estudiantes en la Prueba de Matemáticas. Los estudiantes de colegios no oficiales tienen un desempeño mucho más alto en esta prueba, que aquellos que pertenecen a colegios oficiales, observando grandes diferencias estandarizadas (d > 0,8). No se observaron diferencias importantes en el rendimiento de la prueba según el carácter del colegio: Académico, Técnico o Técnico / Académico. Tampoco teniendo en cuenta el área de ubicación del colegio rural o urbano, puesto que las diferencias estandarizadas no son relevantes. Sin embargo, los estudiantes de colegios con jornada completa superan en alta medida, en el desempeño de la Prueba de Lenguaje, a los estudiantes de colegios con las otras jornadas.

*Tabla 24. **Efecto de las Variables Institucionales en el Desempeño de Lenguaje en las Pruebas Saber 5° del Departamento de Nariño***

Variable institucional		N	Media	DT	d Cohen
Tipo colegio	Femenino	942	318,6	66,3	0,04
	Masculino	3	316,0	78,8	-
	Mixto	23.117	307,1	62,5	0,14

Naturaleza	No Oficial	1.672	355,3	67,6	-
	Oficial	22.390	304,0	60,8	0,84
Carácter	Académico	15.910	310,0	63,7	-
	Técnico	2.271	301,2	58,4	0,14
	Técnico/ Académico	5.881	303,3	61,0	0,11
Área Ubicación	Rural	10.269	300,6	59,7	0,19
	Urbano	13.793	312,7	64,3	-
Jornada	Completa	177	382,8	73,3	-
	Mañana	19.338	307,0	62,1	1,22
	Sabatina	8	309,3	54,2	1,01
	Tarde	3.226	303,3	64,8	1,22
	Única	1.313	315,2	57,2	1,14

Fuente: elaboración propia.

Efecto en la prueba de Matemáticas

En las Pruebas de Matemáticas, como se observa en la Tabla 25, los desempeños de los estudiantes de colegios de tipo masculino resultan mejores que aquellos que pertenecían a colegios femeninos y mixtos que presentan desempeños más bajos que los anteriores con diferencias moderadas respecto a los mismos. Igualmente, los estudiantes de colegios no oficiales obtuvieron mejor rendimiento en Matemáticas que aquellos que pertenecían a colegios oficiales, con diferencias moderadas. No se observó diferencias estandarizadas relevantes en los puntajes de dicha prueba, entre los estudiantes que pertenecían a colegios de carácter Académico, Técnico o Técnico / Académico. Los estudiantes ubicados en el área urbana obtienen resultados similares, que los que corresponden al área rural, con diferencias no relevantes en tamaño del efecto. Finalmente, los estudiantes de colegios con jornada completa superan en alta medida, a los estudiantes de otras jornadas, en el desempeño en la competencia de Matemáticas.

*Tabla 25. **Efecto de las Variables Institucionales en el Desempeño de Matemáticas en las Pruebas Saber 5° del Departamento de Nariño***

Variable Institucional		**N**	**Media**	**DT**	**d Cohen**
Tipo Colegio	Femenino	942	293,3	66,3	0,30
	Masculino	3	313,0	153,3	-
	Mixto	23.117	294,3	66,9	0,28
Naturaleza	No Oficial	1.672	336,3	72,8	-
	Oficial	22.390	291,1	65,3	0,69
Carácter	Académico	15.910	297,2	68,1	-
	Técnico	2.271	287,0	62,4	0,15
	Técnico/ Académico	5.881	289,1	64,5	0,12
Área Ubicación	Rural	10.269	291,2	66,5	0,08
	Urbano	13.793	296,6	67,0	-
Jornada	Completa	177	356,1	75,8	-
	Mañana	19.338	294,8	66,9	0,91
	Sabatina	8	290,8	21,8	0,88
	Tarde	3.226	284,9	66,1	1,07
	Única	1.313	300,7	61,0	0,88

Fuente: elaboración propia.

CAPÍTULO III
MATERIALES Y MÉTODOS

Materiales

Los datos que se utilizaron en esta investigación se obtuvieron del repositorio de datos de acceso libre del Icfes, donde se encuentra almacenada, en archivos planos delimitados, la información socioeconómica, académica e institucional, de los estudiantes de las Instituciones Educativas de Básica Primaria, que presentaron las Pruebas Saber 5º. Los datos más actualizados de este repositorio, al momento de realizar esta investigación, eran de los estudiantes que presentaron las Pruebas Saber 5º en el año 2017. El archivo, consta de 776.436 registros con 56 atributos, cuyos datos se importaron a la tabla *sbr5_estudiantes* de la base de datos *sbr5db* que se creó con el SGBD PostgreSQL.

Para evaluar el desempeño alcanzado en las competencias básicas de las pruebas Saber 5º que presentaron los estudiantes nariñenses en el año 2017, se eliminaron del repositorio *sbr5_estudiantes*, todos los estudiantes que no eran del departamento de Nariño. Como resultado se obtuvo, el conjunto de datos s*br5_narino* con 24062 registros y 56 atributos. El diccionario de datos de la tabla *sbr5_narino* se muestra en la Tabla 26.

Con el fin de descubrir patrones de desempeño académico en las Pruebas Saber 5° de los estudiantes de las Instituciones Educativas del departamento de Nariño, que presentaron la prueba, en el año 2017, se acordó que, si el puntaje del estudiante en la competencia evaluada estaba por encima de la media del puntaje nacional de dicha competencia, se considera su desempeño a nivel de esa competencia bueno y si estaba por debajo de la media nacional, se considera bajo. Esto permite conocer mediante un proceso supervisado, los patrones de aquellos estudiantes

nariñenses cuyo desempeño está tanto sobre la media como bajo la media y centrarse principalmente en los estudiantes cuyo desempeño es bajo.

Tabla 26. Diccionario de Datos del Conjunto de Datos sbr5_narino.

#	Atributo	Descripción	Valores
		Información Personal	
1	ESTU_TIPO-DOCUMENTO	Tipo de Documento	• CC – Cédula de ciudadanía • CCB- Certificado de cabildo • CE – Cédula extranjera • CR – Certificado registraduría • NES-Número establecido por secretaría de educación • NUIP – Número único de identificación personal • NUI – Número único de identificación • PC – Pasaporte colombiano • PE – Pasaporte extranjero • RC – Registro civil de nacimiento • TI – Tarjeta de identidad
2	ESTU_ GENERO	Género	• Hombre • Mujer
3	ESTU_EDAD	Edad del Estudiante	• 9 años o menos • 10 años • 11 años • 12 años o mas
4	ESTU_GRADO	Grado del Estudiante	5º

5	ESTU_FECHA-NACIMIENTO	Fecha de Nacimiento	DD/MM/AAAA
6	PERIODO	Periodo de Presentación	2017
7	ESTU_CONSE-CUTIVO	Id público del Estudiante	Texto
8	ESTU_ ESTUDIANTE	Indica si realizó la inscripción por medio de un colegio (estudiante) o fue individual	ESTUDIANTE
		Información Socioeconómica	
9	FAMI_ EDUCCION PADRE	Nivel Educativo más alto alcanzado por el Padre	• Primaria • Bachillerato • Técnico o Tecnólogo • Universitario - Pregrado • Universitario – Posgrado • NA
10	FAMI_ EDUCACION-MADRE	Nivel Educativo más alto alcanzado por la Madre	• Primaria • Bachillerato • Técnico o Tecnólogo • Universitario - Pregrado • Universitario – Posgrado • NA
11	FAMI_ PERSONAS-HOGAR	Número de personas conforman el hogar donde vive actualmente, incluido el estudiante	2,3,4,5,6
12	FAMI_ CUARTOS HOGAR	Número de cuartos donde duermes las personas del hogar	1,2,3,4,5,6,NA

13	FAMI_ TIENE INTERNET	Si el hogar cuenta con servicio o conexión a internet	No, Sí, NA
14	FAMI_TIENE-SERVICIOTV	Si el hogar cuenta con servicio cerrado de televisión	No, Sí, NA
15	FAMI_TIENE-COMPUTA-DOR	Si en el hogar hay o no computador	No, Sí, NA
16	FAMI_TIENE-LAVADORA	Si en el hogar hay o no lavadora	No, Sí, NA
17	FAMI_TIENE-HORNOMI-CROOGAS	Si en el hogar hay o no microondas	No, Sí, NA
18	FAMI_TIE-NEAUTOMO-VIL	Si en el hogar hay o no automóvil particular	No, Sí, NA
19	FAMI_TIENE-CONSOLAVI-DEOJUEGOS	Si en el hogar hay o no consola de video juegos (PlayStation, Xbox, Nintendo, etc.)	No, Sí, NA
20	FAMI_TRABA-JOLABORPA-DRE	Trabajo que realizó el padre durante la mayor parte del último año	• Trabaja en el hogar; no trabaja o estudia • Es dueño de un negocio pequeño (tiene pocos empleados o no tiene; por ejemplo, tienda, papelería, salón de belleza, lavandería, etc.) • Es dueño de un negocio grande (empresa o institución) o tiene un cargo de nivel directivo o gerencial

20	FAMI_TRABAJOLABORPADRE		• Es dueño de un negocio grande (empresa o institución) o tiene un cargo de nivel directivo o gerencial • Tiene un trabajo de tipo auxiliar administrativo (por ejemplo, secretario o asistente) • Es vendedor, vende por catálogo o trabaja en atención al público • Es agricultor, pescador o jornalero • Trabaja por cuenta propia (por ejemplo, plomero, electricista) • Es operario de máquinas o conduce vehículos (taxista, chofer) • Realiza labores de limpieza, mantenimiento, seguridad o construcción • Trabaja como profesional (por ejemplo, médico, abogado) • Es pensionado • NA
		Información Socioeconómica	
21	FAMI_TRABAJOLABORMADRE	Trabajo que realizó la madre durante la mayor parte del último año	• Trabaja en el hogar; no trabaja o estudia • Es dueño de un negocio pequeño (tiene pocos empleados o no tiene; por ejemplo, tienda, papelería, salón de belleza, lavandería, etc.)

21	FAMI_ TRABAJO LABORMA-DRE		• Es dueño de un negocio grande (empresa o institución) o tiene un cargo de nivel directivo o gerencial • Tiene un trabajo de tipo auxiliar administrativo (por ejemplo, secretario o asistente) • Es vendedor, vende por catálogo o trabaja en atención al público • Es agricultor, pescador o jornalero • Trabaja por cuenta propia (por ejemplo, plomero, electricista) • Es operario de máquinas o conduce vehículos (taxista, chofer) • Realiza labores de limpieza, mantenimiento, seguridad o construcción • Trabaja como profesional (por ejemplo, médico, abogado) • Es pensionado • NA
22	ESTU_ DEDICACION INTERNET	Tiempo al día dedicado a navegar en internet	• No navego en internet • 30 minutos o menos • Entre 30 y 60 minutos • Entre 1 y 3 horas • Más de 3 horas • NA

23	ESTU_COMO-DESPLAZA COLEGIO	Modo de desplazarse al colegio	• Caminando o en bicicleta • En bus o transporte público • En transporte escolar o particular • Otro (burro o caballo, canoa, lancha, etc.) • NA
24	ESTU_ TIEMPO DESPLAZA COLEGIO	Tiempo de desplazamiento de la casa al colegio	• Menos de 15 minutos • Entre 15 y 30 minutos • Entre 30 minutos y una hora • Más de 1 hora NA
25	ESTU_ANOS-PREESCOLAR	Años cursado en Preescolar	• Ninguno • Un año • Dos años • Tres años • No me acuerdo • NA
26	ESTU_ESTE-COLEGIO_ ANOPASADO	Estudió el año pasado en el mismo colegio	No, Sí, NA
27	ESTU_ GRADOCO-MENZO_ES-TECOLEGIO	Desde que grado estudia en el colegio actual	• Primero • Segundo • Tercero • Cuarto • Quinto • NA
28	ESTU_REPRO-BOGRADO	Si ha repetido algún grado	No, Sí, NA
29	ESTU_TAR-DEACLASE	Número de veces que ha llegado tarde el último mes	• Nunca • Una o dos veces por semana • Una o dos veces por mes • Todos los días • NA

30	ESTU_NOFUE-ALCOLEGIO	Número de dias que no ha ido al colegio en el último mes	• Nunca • Uno o dos días • Tres a cinco días • Más de cinco días • NA
31	ESTU_REMU-NERA_OTRA-SACTIVIDADES	Si ha recibido remuneración por actividades diferentes a las escolares	No, Sí, NA
32	ESTU_SERETI-ROCOLEGIO	Tiempo de retiro o suspensión de estudios	• No • Sí, menos de un año • Sí, un año • Sí, dos años o más • NA
33	FAMI_NUMLI-BROS	Numero de libros que hay en el hogar	• 0 a 10 libros • 11 a 25 libros • 26 a 100 libros • Más de 100 libvros • NA

Información del Colegio

34	COLE_COD_ DANE_ESTA-BLECIMIENTO	Código DANE del establecimiento	Texto
35	COLE_NOM-BRE_ESTABLE-CIMIENTO	Nombre del Establecimiento	Texto
36	COLE_GENERO	Indica el género de la población del establecimiento	• FEMENINO • MASCULINO • MIXTO
37	COLE_ NATURALEZA	Indica la naturaleza del establecimiento	• NO OFICIAL • OFICIAL

38	COLE_BILINGUE	Indica si el establecimiento es bilingüe o no	N, S
39	COLE_CARACTER	Indica el carácter del establecimiento	• ACADÉMICO • TÉCNICO • TÉCNICO/ACADÉMICO • NO APLICA
40	COLE_COD_DANE_SEDE	Código DANE de la sede	Texto
41	COLE_NOMBRE_SEDE	Nombre de la Sede	Texto
42	COLE_AREA_UBICACION	Área de ubicación de la sede	• RURAL • URBANO
43	COLE_JORNADA	Jornada de la sede	• COMPLETA • MAÑANA • NOCHE • SABATINA • TARDE • UNICA
44	COLE_COD_MCPIO_UBICACION	Código DANE del municipio donde está ubicada la sede	Número
45	COLE_MCPIO_UBICACION	Nombre del municipio donde está ubicada la sede	Texto
46	COLE_COD_DEPTO_UBICACION	Código DANE del departamento de la sede	Número
47	COLE_DEPTO_UBICACION	Nombre del departamento donde está ubicada la sede	Texto

		Datos Citación del Examen	
48	ESTU_COD_MCPIO_PRESENTACION	Código DANE del municipio presentación del examen	Número
49	ESTU_MCPIO_PRESENTACION	Municipio de presentación del examen	Texto
50	ESTU_COD_DEPTO_PRESENTACION	Código DANE del departamento del municipio de presentación	Número
		Resultados	
51	PUNT_LENGUAJE	Puntaje del estudiante en lenguaje	Rango [100, 500]
52	NIVEL_DESEMP_LENGUAJE	Nivel de desempeño en lenguaje	• Insuficiente • Mínimo • Satisfactorio • Avanzado
53	PUNT_MATEMATICAS	Puntaje del estudiante en matemáticas	Rango [100, 500]
54	NIVEL_DESEMP_MATEMATICAS	Nivel de desempeño en matemáticas	• Insuficiente • Mínimo • Satisfactorio • Avanzado
55	ESTU_INSE_INDIVIDUAL	Índice socioeconómico a nivel de estudiante	Número
56	ESTU_NSE_ESTABLECIMIENTO	Nivel Socioeconómico del Establecimiento	1, 2, 3, 4, 5

Fuente: elaboración propia con datos de Icfes (s.f).

Metodología

Los estudios descriptivos, conciernen y son diseñados para describir, la distribución de variables, sin considerar hipótesis causales o de otra naturaleza. Por esta razón, esta investigación fue de tipo descriptivo, bajo el enfoque cuantitativo, aplicando un diseño no experimental.

Con el fin de llevar a cabo el proceso de minería de datos en forma sistemática y no trivial, se analizaron tres metodologías principales: SEMMA (Azevedo y Santos, 2008), CATALYST (P3TQ) (Pyle, 2003) y CRISP-DM (Chapman, *et al*., 2000), (Shearer, 2016).

SEMMA (Sample (Muestreo), Explore (Exploración), Modify (Modificación), Model (Modelado), Assess (Valoración)), creada por el SAS Institute, se define, como "el proceso de selección, exploración y modelado de grandes volúmenes de datos, para descubrir patrones de negocio desconocidos" SAS Institute (Como se citó en Moine *et al*., 2011, pp. 2-3).

La metodología SEMMA, se encuentra enfocada especialmente, en aspectos técnicos, excluyendo actividades de análisis y comprensión del problema que se está abordando. La metodología Catalyst (Pyle, 2003), conocida como P3TQ (Product, Place, Price, Time, Quantity), plantea la formulación de dos modelos: el Modelo de Negocio y el Modelo de Explotación de Información. El Modelo de Negocio, proporciona una guía de pasos para identificar un problema (o la oportunidad del mismo) y los requerimientos reales de la organización. El Modelo de Explotación de Información, proporciona una guía de pasos para la construcción y ejecución de modelos de minería de datos a partir del Modelo de Negocio. (Moine *et al*., 2011, p. 3).

CRISP–DM (*Cross-Industry Standard Process for Data Mining*), creada por el grupo de empresas SPSS, NCR y Daimer Chrysler en el año 2000, es

actualmente, la guía de referencia más utilizada en el desarrollo de proyectos de minería de datos. Estructura el proceso en seis fases: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación e Implantación Chapman (Como se citó en Rojas Calvo, 2016, p.34). La sucesión de fases, no es necesariamente rígida. Cada fase es descompuesta en varias tareas generales de segundo nivel (Shearer, 2016).

Azevedo y Santos (2008), Moiné *et al*., (2011) y Moiné y Haedo, (2015) comparan las tres metodologías y llegan a la conclusión, de que, aunque se puede establecer un paralelismo claro entre ellas, CRISP-DM es más completo, porque tiene en cuenta la aplicación al entorno de negocio de los resultados y por ello, es la que se adoptó popularmente. En encuestas realizadas en KDNuggets en 2002, 2004, 2007 y 2014, se comprobó que CRISP-DM era la principal metodología utilizada, cuatro veces más que SEMMA (Shearer, 2016) y por esa razón es la metodología que se escogió para realizar esta investigación.

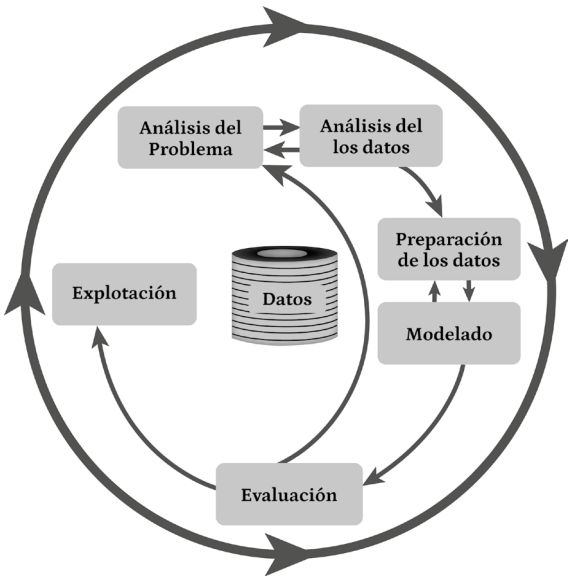
Metodología CRISP-DM

La metodología CRISP-DM (*Cross-Industry Standard Process for Data Mining*) para proyectos de minería de datos, no es la "más actual" o "la mejor", pero es muy útil para comprender esta tecnología o extraer ideas para diseñar o revisar métodos de trabajo para proyectos de similares características. La metodología CRISP-DM, consiste, en un conjunto de tareas que están organizadas en cuatro niveles de abstracción: fases, tareas generales, tareas especializadas e instancias de proceso. Dichos niveles están establecidos, respetando jerarquías en tareas, inician en el nivel más general hasta llegar, finalmente, a los casos más específicos (Chapman, *et al.,* 2000). El modelo, provee una representación completa del ciclo de vida de un proyecto de minería de datos. El proceso es dinámico e iterativo, por lo que la ejecución de los procesos, no es estricta y con frecuencia, se puede pasar de uno a otra, de atrás hacia delante y viceversa. Éstos dependen

del resultado de cada fase o la planeación de la siguiente tarea a ejecutar. Cada fase se estructura en varias tareas generales, las tareas generales se proyectan en tareas específicas, donde finalmente se describen las acciones que deben ser desarrolladas para situaciones definidas (Larose y Larose, 2014).

CRISP-DM está compuesta por seis fases: análisis del problema, análisis de los datos, preparación de los datos, modelado, evaluación y explotación (Gallardo, 2009) como se aprecia en la Figura 11.

Figura 11. Fases de la Metodología CRISP-DM.



Fase de Análisis del Problema

La Fase de Análisis del Problema, se centra en la comprensión de los objetivos y requisitos del proyecto, desde una perspectiva empresarial, y luego, convertir este conocimiento en una definición del problema de minería de datos y un plan preliminar diseñado para alcanzar los objetivos.

Particularmente, en esta fase, se recopiló y seleccionó la información necesaria que permitió a los investigadores conocer y apropiar el conocimiento acerca de las Pruebas Saber 5º y de las competencias que evalúa, haciendo énfasis en Matemáticas y Lenguaje. Este proceso posibilitó la recolección de los datos correctos para obtener resultados adecuados.

Las Pruebas Saber 5°, son una evaluación estandarizada de carácter externo, aplicada periódicamente por el Instituto Colombiano para la Evaluación de la Educación – Icfes, a los estudiantes de Educación Básica Primaria. Esta evaluación, busca identificar el desarrollo de las competencias en áreas fundamentales como las Matemáticas, Competencia Comunicativa en Lenguaje, Competencias Ciudadanas y Ciencias Naturales (Icfes, 2022). De acuerdo con los Lineamientos para las aplicaciones muestral y censal del Icfes (Icfes, 2014b), el diseño de las Pruebas Saber 5° "está alineado con los Estándares Básicos de Competencias, establecidos por el Ministerio de Educación Nacional-MEN, entendidos como referentes comunes, a partir de los cuales, es posible establecer, qué tanto los estudiantes y el sistema educativo en su conjunto, están cumpliendo con unas expectativas de calidad, en términos de lo que saben y saben hacer".

Según la Guía de Orientación de las Pruebas Saber 5º (MEN, 2017), en la Prueba de Matemáticas, se evalúa el uso flexible de las Matemáticas en diversas situaciones. "En las pruebas, se asume la perspectiva integradora de los lineamientos curriculares y los Estándares Básicos de Competencias, como derroteros para la evaluación, respecto a los conocimientos básicos, procesos y contextos" (Icfes, 2022b, p. 35). "Las Competencias Comunicativas en Lenguaje, se evalúan a través de dos instrumentos, uno enfocado en evaluar la comprensión lectora y otro enfocado en evaluar la competencia para escribir de los estudiantes" (Icfes, 2022b, p. 42). "En las Competencias Ciudadanas, se evalúa la capacidad de los estudiantes para participar, en su calidad de ciudadanos, de manera constructiva y activa en la sociedad" (Icfes, 2022b, p. 42); y en la Prueba de Ciencias Naturales,

> Se evalúa la capacidad que tiene el estudiante de comprender y utilizar nociones, conceptos y teorías de las Ciencias Naturales, en la solución de problemas. Evalúa también la habilidad del estudiante, para explicar cómo ocurren algunos fenómenos de la naturaleza, basado en observaciones, patrones y conceptos propios del conocimiento científico. La prueba, además involucra, en la evaluación el proceso de indagación, que incluye observar y relacionar patrones en los datos para derivar conclusiones de fenómenos naturales (Icfes, 2022b, p. 61).

En el caso de esta investigación, se analizaron los resultados en Matemáticas y Lenguaje, competencias que se evaluaron en el año 2017.

Para poder analizar el desempeño de los estudiantes de Básica Primaria del departamento de Nariño en las competencias básicas de las pruebas Saber 5º, con minería de datos, se acordó, mediante el uso de un proceso supervisado que, si el puntaje del estudiante estaba por encima de la media nacional en la competencia evaluada, su desempeño en la competencia era bueno y si estaba por debajo de la media nacional, no era bueno. De esta manera se identificarían patrones de desempeño académico en las pruebas Saber 5º de los estudiantes nariñenses que están tanto sobre la media como aquellos que están bajo la media y, centrarse principalmente, en los estudiantes cuyo desempeño es bajo.

Fase de Análisis de los Datos

La Fase de Análisis de los Datos, comienza con una colección inicial de datos y procesos, con actividades, con el objetivo de familiarizarse con los datos, identificar la calidad de los datos, para descubrir las primeras señales dentro de los datos y detectar temas interesantes, para poder formular hipótesis de información oculta.

En esta fase, se identificó, recopiló y familiarizó, con la información socioeconómica, académica e institucional, disponible en las bases de datos del Icfes, correspondiente a los resultados de los Estudiantes de Educación Básica Primaria del departamento de Nariño, que presentaron las Pruebas Saber 5°, en el año 2017. Como resultado se obtuvo un conjunto de datos, al cual se le denominó s*br5_narino,* con 24062 registros y 56 atributos, cuya descripción se puede mirar en la Tabla 26.

Fase de Preparación de los Datos

La Fase de Preparación de los Datos, cubre todas las actividades para construir el conjunto de datos. Estas tareas, son ejecutadas en múltiples oportunidades y sin orden. Las tareas incluyen selección y transformación de tablas, registros y atributos, y limpieza de datos para las herramientas de modelado.

En esta fase, se realizó inicialmente un análisis de la calidad de los datos del repositorio *sbr5_narino,* con el fin de conocer, por cada atributo, el número de valores distintos, el número de valores nulos, el valor máximo, valor mínimo, moda, media y así determinar, cuáles técnicas de limpieza de datos se podían aplicar (Timarán et al., 2019).

Los 56 atributos del repositorio base, considerados por el Icfes como los más importantes para capturar la información de las Pruebas Saber 5°, fueron depurados teniendo en cuenta la calidad de los datos y las técnicas de minería de datos a aplicar; se limpiaron (eliminación de datos nulos y valores constantes) e integraron los datos, se generaron atributos adicionales a partir de los existentes por ganancia de información, se realizaron transformaciones o cambios de formato a los valores de los atributos que se consideraron necesarios, se eliminaron los atributos reemplazados, así como los atributos con alto porcentaje de valores faltantes (nulos). Con el fin de facilitar la detección de patrones de desempeño académico,

se transformaron los valores numéricos, de ciertos atributos, a valores categóricos, teniendo en cuenta un rango de valores o que las frecuencias por cada valor sean proporcionales, para evitar sesgos al construir los modelos de minería de datos (Timarán et al., 2019).

Como resultado de esta fase, se obtuvo un repositorio de datos limpio y transformado, denominado *sbr5_narino_final,* con 24062 registros y 15 atributos, listos para aplicarles las técnicas de minería de datos. Este conjunto de datos final, sirvió de base para el descubrimiento de patrones de desempeño académico en las Pruebas Saber 5º, de los estudiantes del departamento de Nariño que presentaron esta prueba, en el año 2017.

El diccionario de datos del conjunto de datos *sbr5_narino_final*, se muestra en la Tabla 27.

*Tabla 27. **Diccionario de Datos de los Conjuntos de Datos sbr5_narino_final***

No	ATRIBUTO	DESCRIPCIÓN
1	ESTU_GENERO	Sexo del estudiante
2	ESTU_EDAD	Intervalo de edad al cual pertenece el estudiante en el momento de presentar la prueba
3	ESTU_REPROBOGRADO	Si reprobó o no grado
4	FAMI_EDUCACION_MADRE	Máximo nivel educativo de la madre
5	FAMI_EDUCACION_PADRE	Máximo nivel educativo del padre
6	FAMI_HACINAMIENTO	Índice de Hacinamiento de la familia

7	FAMI_TIC	Condición de uso de TIC en el hogar del estudiante
8	FAMI_ ELECTRODOMESTICOS	Condición relacionada con los electrodomésticos en hogar del estudiante
9	COLE_GENERO	Si el colegio es mixto, femenino, masculino
10	COLE_NATURALEZA	Si el colegio es oficial o privado
11	COLE_CARACTER	Si el colegio es técnico, académico o técnico/académico.
12	COLE_AREA_UBICACIÓN	Si el colegio está en el área urbana o rural
13	COLE_JORNADA	Jornada de estudio del colegio
14	COLE_REGIÓN_ UBICACIÓN	Región del departamento de Nariño en la que se ubica el colegio
15	CLASE_DESEMP_ MATEMATICAS	Desempeño en matemáticas (sobre media, bajo media nacional o departamental)
15	CLASE_DESEMP_ LENGUAJE	Desempeño en lenguaje (sobre media, bajo media nacional o departamental)
15	CLASE_DESEMP_ GENERAL	Desempeño promedio de los estudiantes (sobre media, bajo media nacional o departamental)

Fuente: elaboración propia.

El proceso de limpieza y transformación de este repositorio, se realizó de la siguiente manera: se efectuó una primera selección de atributos y se descartaron aquellos atributos con un alto porcentaje de valores nulos, por la imposibilidad de encontrar sus valores a través de fuentes externas de datos; tampoco se tomaron en cuenta los atributos con valores constantes y aquellos que servían de identificadores de cada estudiante. (Timarán et al., 2021, p. 85).

La alta dimensionalidad, es un problema para el descubrimiento de patrones con minería de datos (Hernández *et al.,* 2005). Uno de los criterios utilizados para resolver este problema, es reducir el número de atributos a analizar, a través de la transformación de estos, en nuevos atributos, que generalicen los datos y que brinden más información. Teniendo en cuenta este criterio, en el repositorio de datos *sbr5_narino_final*, se seleccionaron los atributos que, por sí mismos, no tenían mayor significado, pero que si se consolidaban en un nuevo, adquirían mayor semántica. Por esta razón, se crearon nuevos atributos, en reemplazo de éstos, obteniendo atributos más representativos para el estudio y reduciendo el número de variables a tener en cuenta en la investigación. En la Tabla 28, se detalla este proceso de discretización para el atributo *estu_edad*. (Timarán et al., 2021).

Tabla 28. Valores Discretizados del Atributo estu_edad

Valor	No. Estudiantes
9 años o menos	924
10 años	9690
11 años	7660
12 años o mas	5788

Fuente: esta investigación.

En esta investigación, se consideró conveniente crear un nuevo atributo denominado *cole_region_ubicación,* que hace referencia a la subregión geográfica del departamento de Nariño, donde se encuentra la institución educativa, teniendo en cuenta los municipios del departamento, para el caso de *sbr5_narino_final*. Las subregiones del departamento de Nariño, se muestra en la Tabla 29.

*Tabla 29. **Subregiones de Nariño***

Subregión	**Municipios**
CENTRO	Pasto, Chachagüí, La Florida, Nariño, Tangua, Yacuanquer.
GUAMBUYACO	El Peñol, El Tambo, La Llanada, Los Andes, Sotomayor.
JUANAMBÚ	Arboleda, Buesaco, La Unión, San Pedro de Cartago, San Lorenzo.
LA CORDILLERA	Cumbitara, El Rosario, Leiva, Policarpa, Taminango.
LA SABANA	Guaitarilla, Imués, Ospina, Sapuyes, Túquerres.
LOS ABADES	Providencia, Samaniego, Santacruz.
OBANDO	Aldana, Contadero, Córdoba, Cuaspud, Cumbal, Funes, Iles, Ipiales, Guachucal, Gualmatán, Potosí, Pupiales, Puerres.
OCCIDENTE	Ancuya, Consacá, Linares, Sandoná.
PACÍFICO SUR	Francisco Pizarro, Tumaco.
PIEDEMONTE COSTERO	Mallama, Ricaurte.

RÍO MAYO	Albán, Belén, Colón, El Tablón de Gómez, La Cruz, San Bernardo, San Pablo.
SANQUIANGA	El Charco, La Tola, Mosquera, Olaya Herrera, Santa Bárbara.
TELEMBI	Barbacoas, Magüí Payán, Roberto Payán.

Fuente: Gobernación del departamento de Nariño (s.f.) citado en Martínez (2019).

Se creó un índice para determinar la condición de electrodomésticos con que cuenta el hogar del estudiante. Este índice se denominó *fami_electrodomesticos*. El índice, es el resultado de la sumatoria de los valores de la presencia (1) o ausencia (0) de cada electrodoméstico en la vivienda. Si el índice es 2, entonces la condición de electrodomésticos es buena, si el índice es 1, entonces la condición de electrodomésticos es media, y finalmente, si el índice es 0, entonces la condición de electrodomésticos es mala. En la Tabla 30, se muestran las variables que intervienen en el cálculo del índice *fami_electrodomesticos* (Timarán et al., 2021).

Tabla 30. Cálculo del Índice para fami_electrodomesticos

Electrodomésticos	Si	No
fami_tienelavadora	1	0
fami_tienehornomicroogas	1	0

Fuente: esta investigación.

Se creó el índice *fami_tic* que tiene en cuenta el número de servicios con que cuenta la vivienda. Se procedió de similar manera que para los valores del indice *fami_electrodomesticos*. Si el índice está entre 3 y 4, entonces la condición tic es buena, si el índice esta entre 1 y 2, entonces la condición tic es media y finalmente, si el índice es 0 entonces la condición

tic es mala. En la Tabla 31 se muestran los atributos que intervienen en el cálculo del índice *fami_tic* (Timarán et al., 2021).

Tabla 31. ***Cálculo del Índice para fami_tic***

Servicios	**Si**	**No**
fami_tieneinternet	1	0
fami_tieneserviciotv	1	0
fami_tienecomputador	1	0
fami_ tieneconsolavideojuegos	1	0

Fuente: esta investigación.

Se creó el índice *fami_hacinamiento*. Para asignar los valores a este índice, se calculó el índice de hacinamiento. El hacinamiento, refiere a la relación entre el número de personas que habitan una vivienda o casa y el espacio o número de cuartos disponibles (Spicker *et al.*, 2009). Generalmente, se aceptan los siguientes valores: hasta 2.4 - sin hacinamiento; de 2.5 a 4.9 - hacinamiento medio y de 5.0 o más - hacinamiento crítico.

Teniendo en cuenta estos conceptos, el índice *fami_hacinamiento* para cada estudiante, se obtuvo dividiendo los valores de los atributos *fami_personashogar* entre fami_*cuartoshogar*; variables que fueron reemplazadas por el atributo *fami_hacinamiento*. Los valores de este nuevo atributo se asignaron teniendo en cuenta los aceptados para el hacinamiento (Timarán et al., 2021).

Fase de Modelado

En la Fase de Modelado, se seleccionan y aplican varias técnicas de modelado y se ajustan los parámetros para obtener óptimos resultados.

Hay varias técnicas que tienen requerimientos específicos para la forma de los datos, por lo que, frecuentemente, es necesario volver a la fase de preparación de datos.

En esta fase, se seleccionó un modelo de clasificación por árboles de decisión, como la técnica de minería de datos más adecuada para solucionar el problema del estudio, dada su facilidad y simplicidad para interpretar los patrones obtenidos (Azevedo y Santos, 2008; Hernández *et al*., 2005; Timarán *et al.,* 2017), siendo este un factor decisivo para su aplicación.

La clasificación por árboles de decisión considera clases disjuntas, de forma que el árbol conducirá a una y sólo una hoja, asignando una única clase a la predicción (Hernández y Lorente, 2009). Esta técnica tiene varias ventajas. En primer lugar, el proceso de razonamiento detrás del modelo, resulta claramente evidente, cuando se examina el árbol. Las personas son capaces de comprender los modelos de árboles de decisión después de una breve explicación, por eso se dice que los árboles de decisión son un modelo de caja blanca. Esto contrasta con otras técnicas de modelado de caja negra, en las que la lógica interna puede resultar difícil de averiguar. En segundo lugar, de manera automática, el proceso incluye en su regla únicamente, los atributos que realmente importan en la toma decisiones; los atributos que no contribuyan a la precisión del árbol se omiten (Han *et al*., 2011, Sattler y Dunemann, 2001; Timarán y Millán, 2006). Por otra parte, los árboles de decisión manejan tanto datos numéricos como categóricos y requieren poca preparación de los datos, al contrario de otras técnicas que necesitan la normalización de los datos y manejan un solo tipo de datos. Además, los árboles de decisión pueden ser fácilmente validados utilizando las métricas obtenidas en la matriz de confusión y de ahí determinar la fiabilidad del modelo. Finalmente, los árboles de decisión son robustos y funcionan bien con grandes volúmenes de datos.

Fase de Evaluación

En la Fase de Evaluación, se evalúa el modelo construido, teniendo en cuenta el cumplimiento de los criterios de éxito y se interpretan los resultados.

En esta fase, se estimó el coste del clasificador, para los repositorios de entrenamiento y prueba a través de una matriz de confusión (Sattler y Dunemann, 2001). Se evaluaron los patrones descubiertos, con el fin de determinar su validez, remover los patrones redundantes o irrelevantes y traducir los patrones útiles en términos que sean entendibles para el usuario. La evaluación e interpretación de los patrones descubiertos, se describe en el Capítulo de Resultados y Discusión (Timarán et al., 2021, p. 70).

Fase de Explotación

En la Fase de Explotación o Implementación, el conocimiento obtenido, se transforma en acciones dentro del proceso de negocio.

En esta fase, a través de la difusión de los informes de esta investigación, el conocimiento descubierto, se incorporará al existente y se podrá integrar a los procesos de toma de decisiones del MEN, la Secretaría de Educación Departamental y las directivas de las instituciones educativas de Básica Primaria nariñenses, que velan por la calidad de la Educación Básica Primaria en el departamento de Nariño. Una vez estas instituciones intervengan los factores asociados al desempeño académico en las Pruebas Saber 5°, será posible analizar los resultados y determinar sus efectos.

CAPÍTULO IV
RESULTADOS E INTERPRETACIÓN DE PATRONES

Se seleccionó la tarea de clasificación, con árboles de decisión, como la técnica de minería de datos más adecuada para descubrir patrones asociados al desempeño académico de los estudiantes de Grado Quinto de Educación Primaria en las Pruebas Saber 5°. El modelo de clasificación basado en árboles de decisión, es probablemente el más utilizado y popular, por su simplicidad y facilidad para entender (Han y Kamber, 2001; Sattler y Dunemann, 2001; Timarán y Millán, 2006). La importancia de los árboles de decisión se debe a su capacidad de construir modelos interpretables, siendo este un factor decisivo para su aplicación.

Con esta técnica se pretende, obtener mediante un proceso de aprendizaje supervisado, un modelo que permita predecir para los nuevos casos de estudiantes, cuáles son los factores socioeconómicos, académicos e institucionales asociados al buen (por encima de la media) o bajo (por debajo de la media) desempeño académico en las Pruebas Saber 5°, teniendo en cuenta el puntaje general, el puntaje en Lenguaje y el puntaje en Matemáticas, obtenidos por cada estudiante nariñense en las Pruebas Saber 5°, como atributos clase.

Para esta investigación interesa que el modelo clasifique mejor (mayor precisión) a los estudiantes cuyo desempeño en las pruebas Saber 5º está bajo la media (minimizar los falsos positivos), con el fin de que los patrones que se descubran ayuden para proponer planes de mejoramiento para aquellos estudiantes nariñenses nuevos que estén en riesgo de estar bajo la media, en alguna competencia.

Definición del Modelo

Con el fin de seleccionar la técnica de árboles de decisión que mejor clasificara el conjunto de datos *sbr5_narino_final*, se evaluaron diferentes algoritmos de árboles de decisión utilizando la herramienta WEKA (Witten *et al.,* 2011; Hall *et al*., 2011), con la opción de prueba *cross_validation* con 10 *folds* y con los parámetros por defecto. El resultado de esta evaluación, se muestra en la Tabla 32.

*Tabla 32. **Evaluación Algoritmos de Árboles de Decisión***

Algoritmo	Exactitud	Precisión	Recall	ROC
ADTree	63.46	63.6	63.4	0.68
J48	64.92	64.9	64.9	0.68
Random Forest	64.73	64.7	64.7	0.70
RandomTree	61.95	61.9	62.0	0.64
REPTree	64.27	64.2	64.3	0.68
AdaBoostM1 con J48	63.15	63.1	63.2	0.67
Baggin con J48	64.30	64.2	64.3	0.69
Stacking con J48	53.0	53.0	53.0	0.50
Vote	53.0	53.0	53.0	0.50

Fuente: elaboración propia.

Teniendo en cuenta los parámetros de exactitud, precisión, recall y curva ROC de la Tabla 32, el algoritmo de árboles de decisión que mejor desempeño tuvo, fue el J48 o C4.5. Por esta razón, se escogió este algoritmo para construir el modelo que permita descubrir los patrones de desempeño académico de los estudiantes de Grado Quinto de Primaria, que presentaron las Pruebas Saber 5º.

El algoritmo J48, que implementa al algoritmo C.45 (Quinlan, 1993), se basa en la utilización del criterio de proporción de ganancia de información (*gain ratio*), que evita que las variables con mayor número de posibles valores, salgan beneficiadas en la selección. Además, el algoritmo incorpora una poda del árbol de clasificación, una vez que éste ha sido inducido (Hernández y Lorente, 2009). Uno de los parámetros que se tuvo en cuenta para la poda, fue el factor de confianza C (*confidence level)*, que influye en el tamaño y capacidad de predicción del árbol construido. Cuanto más baja se haga esa probabilidad, se exigirá que la diferencia en los errores de predicción antes y después de podar, sea más significativa para no podar. El valor por defecto de este factor, es del 25%, y conforme va bajando este valor, se permiten más operaciones de poda y, por lo tanto, llegar a árboles cada vez más pequeños (Molina, y García, 2006).

El otro parámetro utilizado para variar el tamaño del árbol, fue el factor M, que especifica el mínimo número de instancias o registros por nodo del árbol (Witten *et al.,* 2011).

Ajuste de Parámetros

Para escoger los mejores parámetros de poda para el árbol de decisión con el algoritmo J48, se construyeron diferentes árboles, variando los parámetros de poda, combinando los valores de C de 25%, 15% y 5% y de M de 1% y 0.5% del número total de registros del conjunto de datos *sbr5_narino_final* (24062) con la opción de prueba *cross_validation* con 10 *folds* y se evaluó el desempeño del algoritmo. En la Tabla 33, se muestra los resultados.

Tabla 33. Ajuste de Parámetros de Poda para sbr5_narino_final

Árbol	C%	M%	Exactitud	Hojas	Profundidad
Tree_CO25M241	25	1	66.24	61	78
Tree_CO25M121	25	0.5	67.25	78	109
Tree_CO15M241	15	1	66.07	45	51
Tree_CO15M121	15	0.5	67.04	63	87
Tree_COO5M241	5	1	65.63	27	35
Tree_COO5M121	5	1	66.42	43	58

Fuente. Elaboración propia.

Como se puede observar en la Tabla 33, la mejor exactitud la tuvo el árbol *Tree_CO25M121* con un 67.25% de instancias correctamente clasificadas, con un factor de confianza C de 25% y un M de 0.5%. Estos serán los parámetros, con los cuales, se construirán los diferentes modelos para identificar los patrones de desempeño académico en las Pruebas Saber 5°, en las Competencias de Lenguaje, Matemáticas y desempeño general de los estudiantes del departamento de Nariño.

Teniendo en cuenta que, para entrenar y probar un modelo de clasificación, se divide los datos en dos conjuntos: entrenamiento y prueba (Witten *et al.,* 2011), se escogió el método de proveer un conjunto de datos de prueba (*Supplied test set*) como el más adecuado para esta investigación.

Utilizando el filtro *Reesampling* de Weka, el conjunto de datos *sbr5_narino_final* se particionó de manera aleatoria y sin reemplazo en seis subconjuntos: dos subconjuntos con 80-20 (80% de entrenamiento y 20% de prueba), dos subconjuntos con 90-10 (90% de entrenamiento y 10% de

prueba) y dos subconjuntos con 70-30 (70% de entrenamiento y 30% de prueba). Se evaluó su desempeño con los parámetros seleccionados C=15% y M=0.5%. El resultado se muestra en la Tabla 34.

Tabla 34. Particionamiento del Conjunto de Datos sbr5_narino_final

Árbol	No. registros Entrenamiento	No. Registros Prueba	Exactitud
Tree_ C025M121_70_30	16,843	7,219	66.51
Tree_ C025M121_80-20	19,249	4,813	65.67
Tree_ C025M121_90-10	21,655	2,407	64.60

Fuente. Elaboración propia.

De acuerdo a los resultados de la Tabla 55, se escogieron los parámetros para el particionamiento del conjunto de datos *sbr5_narino_final*, el 70% para entrenamiento y 30% para prueba, por su mayor exactitud. Con estos parámetros se construyeron los diversos modelos de árboles de decisión para la detección de patrones en las competencias de lenguaje, matemáticas y desempeño general en las pruebas Saber 5º.

Construcción de Modelos

Teniendo en cuenta los parámetros de evaluación anteriores, se procedió a construir los diferentes árboles de decisión con el algoritmo J48. Por ser los árboles de decisión un modelo de aprendizaje supervisado, se transformaron los puntajes obtenidos en cada competencia evaluada

en el Saber 5º en el 2017, así como el puntaje general de cada estudiante nariñense a variables categóricas, con los valores "sobre la media ", y "bajo la media" con relación a la media nacional obtenida por todos los estudiantes de Quinto de Primaria que presentaron las pruebas Saber 5º en el año 2017. Este atributo se escogió como la variable clase o '*target*'. En la Tabla 35, se puede mirar la distribución de los estudiantes observados para cada clase a nivel del departamento de Nariño, en cada competencia.

Tabla 35. ***Distribución por Clases de sbr5_narino_final***

Clase	Bajo la Media	Sobre la Media	Total
Punt._General	13,559	10,503	24,062
Punt._Lenguaje	13,151	10,911	24,062
Punt._Matemáticas	13,499	10,563	24,062

Fuente: elaboración propia.

Descubrimiento de Patrones de Desempeño Académico en la Competencia de Lenguaje en las Pruebas Saber 5º de los Estudiantes del Departamento de Nariño

Para el Descubrimiento de Patrones de Desempeño Académico en la Competencia de Lenguaje en las Pruebas Saber 5º presentada por los estudiantes de Grado Quinto de Primaria del departamento de Nariño en el año 2017, se tomó el conjunto de datos *sbr5_narino_final* particionado en 70% para entrenamiento (16,843 registros) y 30% para prueba (7,219 registros). Se tomó como variable clase u objetivo, el atributo denominado *CLASE_DESEMP_LENGUAJE* (Tabla 27) con los valores sobre la media y bajo la media. En el conjunto de entrenamiento, se observa 7,610 casos de estudiantes que están sobre la media y 9,233 casos de estudiantes, que están bajo la media. Como parámetros de poda del árbol de decisión, se tomaron los valores para C=25% y para M=0.5%, según el proceso de ajuste de parámetros anteriormente descrito.

Con estos parámetros, se procedió a construir el modelo con el algoritmo J48. En la Figura 12, se muestra la Matriz de Confusión, con los parámetros de evaluación del modelo y en la Figura 12 se muestra el árbol de decisión obtenido para la Competencia de Lenguaje en las Pruebas Saber 5°, una vez aplicado un proceso de pospoda según los parámetros de soporte mínimo y confianza mínima. Para este caso se tuvo en cuenta solo aquellos ramales del árbol que superen un mínimo soporte de 0.5 % (i.e. mínimo número de ejemplos que llegan a un nodo) y una confianza mínima de 60% (i.e. el mínimo de ejemplos correctamente clasificados en cada nodo).

Figura 12. Matriz de Confusión del árbol para la competencia de Lenguaje

```
=== Summary ===

Correctly Classified Instances        4674               64.7458 %
Incorrectly Classified Instances      2545               35.2542 %
Kappa statistic                          0.2885
Mean absolute error                      0.4184
Root mean squared error                  0.4571
Relative absolute error                 84.3659 %
Root relative squared error             91.7554 %
Total Number of Instances             7219

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0,685    0,397    0,672     0,685   0,678      0,289    0,721     0,752     BAJO MEDIA
                 0,603    0,315    0,617     0,603   0,610      0,289    0,721     0,650     SOBRE MEDIA
Weighted Avg.    0,647    0,360    0,647     0,647   0,647      0,289    0,721     0,705

=== Confusion Matrix ===

   a   b   <-- classified as
2683 1235 |   a = BAJO MEDIA
1310 1991 |   b = SOBRE MEDIA
```

Figura 13. Árbol podado ***para la competencia de Lenguaje del Saber 5º de Nariño***.

```
J48 pruned tree
------------------
cole_naturaleza = OFICIAL
|   cole_region_ubicacion = LOS ABADES
|   cole_region_ubicacion = LA CORDILLERA
|   |   fami_educacionmadre = Bachillerato: BAJO MEDIA (211.0/85.0)
|   cole_region_ubicacion = OBANDO
|   |   estu_reprobogrado = Si
|   |   |   cole_caracter = ACADEMICO
|   |   |   cole_caracter = TECNICO/ACADEMICO: BAJO MEDIA (136.0/54.0)
|   |   estu_reprobogrado = No
|   |   |   fami_educacionmadre = Bachillerato
|   |   |   |   estu_edad = 10 anos
|   |   |   |   |   cole_caracter = TECNICO/ACADEMICO: SOBRE MEDIA (152.0/59.0)
|   |   |   fami_educacionmadre = Universitario - Posgrado: SOBRE MEDIA (143.0/53.0)
|   cole_region_ubicacion = PASTO
|   |   estu_reprobogrado = Si
|   |   |   fami_educacionpadre = Bachillerato: BAJO MEDIA (354.0/143.0)
|   |   estu_reprobogrado = No
|   |   |   cole_area_ubicacion = URBANO
|   |   |   |   estu_edad = 11 anos: SOBRE MEDIA (784.0/268.0)
|   |   |   |   estu_edad = 10 anos: SOBRE MEDIA (1181.0/432.0)
|   cole_region_ubicacion = JUANAMBU
|   |   estu_reprobogrado = Si: BAJO MEDIA (240.0/82.0)
|   cole_region_ubicacion = PACIFICO SUR: BAJO MEDIA (2688.0/392.0)
|   cole_region_ubicacion = CENTRO
|   |   estu_genero = Mujer: SOBRE MEDIA (184.0/70.0)
|   cole_region_ubicacion = TELEMBI: BAJO MEDIA (731.0/125.0)
|   cole_region_ubicacion = RIO MAYO
|   |   cole_caracter = ACADEMICO: SOBRE MEDIA (469.0/169.0)
|   |   cole_caracter = TECNICO: BAJO MEDIA (127.0/44.0)
|   cole_region_ubicacion = SANQUIANGA: BAJO MEDIA (913.0/202.0)
cole_naturaleza = NO OFICIAL
|   cole_region_ubicacion = OBANDO: SOBRE MEDIA (209.0/79.0)
|   cole_region_ubicacion = PASTO: SOBRE MEDIA (789.0/158.0)

Number of Leaves  :     80

Size of the tree :      109
```

Analizando los resultados obtenidos en el modelo que se muestra en la Figura 13, sobre el desempeño académico en la Competencia de Lenguaje de los Estudiantes de Grado Quinto de las instituciones educativas de Básica Primaria de Nariño, que presentaron las Pruebas Saber 5º en el año 2017,

se puede observar, que este clasifica correctamente, al 64.74% del total de instancias probadas y clasifica incorrectamente al 35.26%.

Evaluando el modelo con la Matriz de Confusión de la Figura 12, se observó que, de 7,219 casos de estudiantes probados, el 54.3 % (3,918 instancias) están bajo la media y 45.7% (3,301 instancias) están sobre la media y el modelo predice al 55.3% (3,993 instancias) bajo la media y el 44.7% (3,226) sobre la media.

Por otra parte, el modelo predice correctamente a 2,683 casos de estudiantes nariñenses cuyo desempeño en lenguaje están bajo la media (verdaderos negativos -TN) y a 1,991 casos de estudiantes nariñenses que están sobre la media (verdaderos positivos -TP). Además, 1,235 casos cuyo desempeño estaba bajo la media, el modelo los clasifica incorrectamente, como sobre la media (falsos positivos-FP) y 1,310 casos cuyo desempeño estaba sobre la media, el modelo los clasifica incorrectamente, como bajo la media (falsos negativos-FN). De acuerdo a esta matriz, el número de falsos positivos que el modelo clasifica es menor que el número de falsos negativos, lo que significa que para este modelo es menos costoso reconocer a aquellos estudiantes nariñenses cuyo desempeño en lenguaje en las pruebas Saber 5º está bajo la media y que, para esta investigación, es el tipo de estudiantes que interesa.

En cuanto a las métricas de evaluación del modelo de clasificación, para el caso de los estudiantes nariñenses que están bajo la media en el desempeño en lenguaje, el modelo tiene una precisión de predicción de 0.672, lo cual quiere decir que, del total de casos predichos que están bajo la media, el 67,2% están correctos. La sensibilidad (TPR) y exhaustividad (*Recall*) del modelo, es de 0.685, lo cual indica que, el modelo clasifica correctamente al 68.5 % de los estudiantes que realmente están bajo la media. Por otra parte, la tasa de falsos negativos (*FP Rate*) del modelo es de 0.397, lo que significa, que el 39.7% de estudiantes que estaban sobre

la media fueron clasificados como bajo la media. La medida F (*F-Measure* o *F1-Score*) es de 0.678 lo que significa que, la media armónica entre la precisión y la exhaustividad de los que están bajo la media es buena, ya que es mayor que 0.5 y más cercano a 1.

Para el caso de los estudiantes nariñenses que están sobre la media en el desempeño en Lenguaje, el modelo tiene una precisión de predicción de 0.617, lo que quiere decir que, del total de casos predichos que están sobre la media, el 61,7% están correctos. La sensibilidad (TPR) y exhaustividad (*Recall*) del modelo, es de 0.603, lo cual indica que, el modelo clasifica correctamente al 60.3 % de los estudiantes que realmente están sobre la media. Por otra parte, la tasa de Falsos Positivos (*FP Rate*) del modelo es de 0.315, lo que significa, que el 31.5% de estudiantes que estaban bajo la media fueron clasificados como sobre la media. La medida F (*F-Measure* o *F1-Score*) es de 0.610 lo que significa que la media armónica entre la precisión y la exhaustividad de los que están sobre la media no es tan buena como la de los casos que están bajo la media, pero es mayor que 0.5.

El área bajo la curva ROC (*ROC Area*) del modelo es de 0.721(mayor que 0.5) lo que indica, que el modelo tiene una alta probabilidad de que sea capaz de distinguir los valores de la clase sobre la media de los valores de la clase bajo la media, ya que es capaz de detectar más verdaderos positivos (TP) y verdaderos negativos (TN) que de falsos negativos (FN) y falsos positivos (FP).

En la combinación de estas métricas, se aprecia un mejor desempeño del modelo para los que están bajo la media, que como se expresó antes, es el tipo de estudiantes que interesan para la toma de decisiones.

Para escoger los patrones más representativos del árbol obtenido (Figura 13), se tuvo en cuenta, aquellos que superen un mínimo soporte de 0.5 % (i.e. mínimo número de ejemplos que llegan a un nodo) y una

confianza mínima de 60% (i.e. el mínimo de ejemplos correctamente clasificados en cada nodo) con relación al total de ejemplos del conjunto de entrenamiento (16,843). Entre los patrones más importantes según su soporte, están:

Patrón 1. Si el estudiante asiste a un colegio oficial y está ubicado en la Subregión Pacífico Sur de Nariño, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º***, tiene mayor probabilidad de estar bajo la media nacional. El 16% del total de estudiantes evaluados se clasifican de esta manera. El 85.4 % de los estudiantes que tienen estas características, están correctamente clasificados y el 24.9% del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 2. Si el estudiante asiste a un colegio oficial, ubicado en la Subregión Pasto, de Nariño, no ha reprobado grados, el colegio está situado en el área urbana y pertenece al grupo etario de los 10 años, entonces su desempeño académico en la competencia de lenguaje de las pruebas Saber 5° tiene mayor probabilidad de estar sobre la media nacional El 7% del total de estudiantes evaluados se clasifican de esta manera. El 63.4 % de los estudiantes que tienen estas características, están correctamente clasificados y el 9.8% del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 3. Si el estudiante asiste a un colegio oficial y está ubicado en la subregión Sanquianga de Nariño, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º***, tiene mayor probabilidad de estar bajo la media nacional. El 5.4% del total de estudiantes evaluados se clasifican de esta manera. El 77.9 % de los estudiantes que tienen estas características, están correctamente clasificados y el 7.7% del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 4. Si el estudiante asiste a un colegio no oficial y está ubicado en la Subregión Pasto de Nariño, entonces su desempeño académico en

la Competencia de Lenguaje del Saber ***5º***, tiene mayor probabilidad de estar sobre la media nacional. El 4.7% del total de estudiantes evaluados se clasifican de esta manera. El 80 % de los estudiantes que tienen estas características, están correctamente clasificados y el 8.3% del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 5. Si el estudiante asiste a un colegio oficial, ubicado en la Subregión Pasto de Nariño, no ha reprobado grados, el colegio está situado en el área urbana y pertenece al grupo etario de los 11 años entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º***, tiene mayor probabilidad de estar sobre la media nacional. El 4.7% del total de estudiantes evaluados se clasifican de esta manera. El 65.8 % de los estudiantes que tienen estas características, están correctamente clasificados y el 6.8% del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 6. Si el estudiante asiste a un colegio oficial y está ubicado en la Subregión Telembí de Nariño, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º***, tiene mayor probabilidad de estar bajo la media nacional. El 4.3% del total de estudiantes evaluados se clasifican de esta manera. El 82.9 % de los estudiantes que tienen estas características, están correctamente clasificados y el 6.6% del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 7. Si el estudiante asiste a un colegio oficial, ubicado en la subregión Rio Mayo de Nariño y el colegio es de carácter académico entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º***, tiene mayor probabilidad de estar sobre la media nacional. El 2.8% del total de estudiantes evaluados se clasifican de esta manera. El 64 % de los estudiantes que tienen estas características, están correctamente clasificados y el 3.9% del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 8. Si el estudiante asiste a un colegio oficial, ubicado en la Subregión Pasto de Nariño, ha reprobado grados y el máximo nivel educativo del padre es bachillerato, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º*** tiene mayor probabilidad de estar bajo la media nacional. El 2.1% del total de estudiantes evaluados, se clasifican de esta manera. El 60 % de los estudiantes que tienen estas características, están correctamente clasificados y el 2.3% del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 9. Si el estudiante asiste a un colegio oficial, ubicado en la subregión Juanambú de Nariño y ha reprobado grados, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º***, tiene mayor probabilidad de estar bajo la media nacional. El 1.4% del total de estudiantes evaluados se clasifican de esta manera. El 65.8 % de los estudiantes que tienen estas características, están correctamente clasificados y el 1.7% del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 10. Si el estudiante asiste a un colegio oficial, ubicado en la subregión La Cordillera de Nariño y el máximo nivel educativo de la madre es Bachillerato, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º*** tiene mayor probabilidad de estar bajo la media nacional. El 1.3% del total de estudiantes evaluados se clasifican de esta manera. El 60% de los estudiantes que tienen estas características, están correctamente clasificados y el 1.4% del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 11. Si el estudiante asiste a un colegio no oficial y está ubicado en la subregión Obando de Nariño, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º*** tiene mayor probabilidad de estar sobre la media nacional. El 1.2% del total de estudiantes evaluados se clasifican de esta manera. El 62.2 % de los estudiantes que tienen estas

características, están correctamente clasificados y el 1.7% del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 12. Si el estudiante asiste a un colegio oficial, ubicado en la subregión Centro de Nariño y es de sexo femenino, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º*** tiene mayor probabilidad de estar sobre la media nacional. El 1.1% del total de estudiantes evaluados, se clasifican de esta manera. El 62 % de los estudiantes que tienen estas características, están correctamente clasificados y el 1.5% del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 13. Si el estudiante asiste a un colegio oficial, ubicado en la subregión Obando de Nariño, no ha reprobado grados, el máximo nivel educativo de la madre es bachillerato, pertenece al grupo etario de los 10 años y el colegio es de carácter técnico/académico, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º*** tiene mayor probabilidad de estar sobre la media nacional. El 0.9% del total de estudiantes evaluados se clasifican de esta manera. El 61.2% de los estudiantes que tienen estas características, están correctamente clasificados y el 1.2% del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 14. Si el estudiante asiste a un colegio oficial, ubicado en la Subregión Rio Mayo de Nariño y el colegio es de carácter técnico, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º*** tiene mayor probabilidad de estar bajo la media nacional. El 0.8% del total de estudiantes evaluados, se clasifican de esta manera. El 65.4 % de los estudiantes que tienen estas características, están correctamente clasificados y el 0.9% del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 15. Si el estudiante asiste a un colegio oficial, ubicado en la Subregión Obando de Nariño, no ha reprobado grados y el máximo nivel educativo de la madre es universitario-posgrado, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º*** tiene mayor probabilidad de estar sobre la media nacional. El 0.8% del total de estudiantes evaluados, se clasifican de esta manera. El 62.9% de los estudiantes que tienen estas características, están correctamente clasificados y el 1.2% del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 16. Si el estudiante asiste a un colegio oficial, ubicado en la Subregión Obando de Nariño, ha reprobado grados y el colegio es de carácter técnico/académico, entonces su desempeño académico en la Competencia de Lenguaje del Saber ***5º***, tiene mayor probabilidad de estar bajo la media nacional. El 0.8% del total de estudiantes evaluados se clasifican de esta manera. El 60.3% de los estudiantes que tienen estas características, están correctamente clasificados y el 0.9% del total de estudiantes observados que están bajo la media, cumplen este patrón.

Descubrimiento de Patrones de Desempeño Académico en la Competencia de Matemáticas en las Pruebas Saber 5° de los Estudiantes del Departamento de Nariño

Para el Descubrimiento de Patrones de Desempeño Académico en la Competencia de Matemáticas en las Pruebas Saber 5°, presentada por los Estudiantes de Grado Quinto de Primaria del departamento de Nariño en el año 2017, se tomó el conjunto de datos *sbr5_narino_final* particionado en 70% para entrenamiento (16,843 registros) y 30% para prueba (7,2019 registros). Se tomó como variable clase u objetivo, el atributo denominado *CLASE_DESEMP_MATEMATICAS* (Tabla 43), con los valores sobre la media y bajo la media. En el conjunto de entrenamiento, se observa 7,360 casos de estudiantes que están sobre la media y 9,493 casos de estudiantes que

están bajo la media. Como parámetros de poda del árbol de decisión, se tomaron los valores para C=25% y para M=0.5%, según el proceso de ajuste de parámetros anteriormente descrito.

Con estos parámetros, se procedió a construir el modelo con el algoritmo J48. En la Figura 14, se muestra la Matriz de Confusión, con los parámetros de evaluación del modelo y en la Figura 15, se muestra el árbol de decisión obtenido por los estudiantes nariñenses para la Competencia de Matemáticas en las Pruebas Saber 5º, una vez aplicado un proceso de pospoda según los parámetros de soporte y confianza mínimos. Para este caso se tuvo en cuenta solo aquellos ramales del árbol que superen un mínimo soporte de 0.5 % (i.e. mínimo número de ejemplos que llegan a un nodo) y una confianza mínima de 60% (i.e. el mínimo de ejemplos correctamente clasificados en cada nodo).

Figura 14. Matriz de confusión para el árbol de la competencia de Matemáticas.

```
=== Summary ===

Correctly Classified Instances        4709               65.2306 %
Incorrectly Classified Instances      2510               34.7694 %
Kappa statistic                          0.3047
Mean absolute error                      0.4228
Root mean squared error                  0.4598
Relative absolute error                 85.76   %
Root relative squared error             92.5026 %
Total Number of Instances             7219

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0,637    0,328    0,707     0,637    0,670      0,307    0,709     0,748     BAJO MEDIA
                 0,672    0,363    0,597     0,672    0,632      0,307    0,709     0,621     SOBRE MEDIA
Weighted Avg.    0,652    0,344    0,658     0,652    0,653      0,307    0,709     0,692

=== Confusion Matrix ===

   a   b   <-- classified as
2551 1455 |   a = BAJO MEDIA
1055 2158 |   b = SOBRE MEDIA
```

Analizando los resultados obtenidos en el modelo que se muestra en la Figura 14 sobre el desempeño académico en la Competencia de Matemáticas de los Estudiantes de Grado Quinto de las Instituciones Educativas de Básica Primaria de Nariño, que presentaron las Pruebas Saber 5º en el año 2017, se puede observar, que este clasifica correctamente al 65.23% del total de instancias probadas y clasifica incorrectamente al 34.77%.

Evaluando el modelo con la Matriz de Confusión de la Figura 14, se observó que, de 7,219 casos de estudiantes probados, el 55.5 % (4,006 instancias) están bajo la media y 44.5% (3,213 instancias) están sobre la media y el modelo predice al 50% (3,606 instancias) bajo la media y el 50% (3,613) sobre la media.

Por otra parte, el modelo predice correctamente a 2,551 casos de estudiantes nariñenses cuyo desempeño en matemáticas están bajo la media (verdaderos negativos -TN) y a 2,158 casos de estudiantes nariñenses que están sobre la media (verdaderos positivos -TP). Además, 1,445 casos, cuyo desempeño estaba bajo la media, el modelo los clasifica incorrectamente como sobre la media (falsos positivos-FP) y 1,055 casos, cuyo desempeño estaba sobre la media, el modelo los clasifica incorrectamente como bajo la media (falsos negativos-FN). De acuerdo a esta matriz, el número de falsos positivos que el modelo clasifica es mayor que el número de falsos negativos, lo que significa que para este modelo es más costoso reconocer a aquellos estudiantes nariñenses cuyo desempeño en matemáticas en las pruebas Saber 5º está bajo la media y que, para esta investigación, es el tipo de estudiantes que interesa.

En cuanto a las métricas de evaluación del modelo de clasificación, para el caso de los estudiantes nariñenses, que están bajo la media en el desempeño en Matemáticas, el modelo tiene una precisión de predicción de 0.707, lo que quiere decir que, del total de casos predichos que están bajo la media, el 70,7% están correctos. La sensibilidad (TPR) y exhaustividad

(*Recall*) del modelo, es de 0.637, lo que indica, que el modelo clasifica correctamente al 63.7 % de los estudiantes que realmente están bajo la media. Por otra parte, la tasa de Falsos Negativos (*FP Rate*) del modelo es de 0.328, lo que significa que el 32.8% de estudiantes que estaban sobre la media fueron clasificados como bajo la media. La medida F (*F-Measure* o *F1-Score*) es de 0.670 lo que significa que la media armónica entre la precisión y la exhaustividad de los que están bajo la media es buena, ya que es mayor que 0.5 y más cercano a 1.

Figura 15. **Árbol** ***podado para la competencia de Matemáticas del Saber 5º de Nariño.***

```
J48 pruned tree
------------------

cole_naturaleza = OFICIAL
|   cole_region_ubicacion = LOS ABADES
|   |   cole_area_ubicacion = RURAL: SOBRE MEDIA (241.0/92.0)
|   cole_region_ubicacion = PASTO
|   |   estu_reprobogrado = Si: BAJO MEDIA (693.0/264.0)
|   cole_region_ubicacion = JUANAMBU
|   |   estu_reprobogrado = Si: BAJO MEDIA (240.0/96.0)
|   cole_region_ubicacion = PACIFICO SUR: BAJO MEDIA (2688.0/353.0)
|   cole_region_ubicacion = CENTRO
|   |   estu_edad = 10 anos: SOBRE MEDIA (209.0/84.0)
|   cole_region_ubicacion = OCCIDENTE
|   |   fami_educacionmadre = Bachillerato: SOBRE MEDIA (218.0/101.0)
|   cole_region_ubicacion = TELEMBI: BAJO MEDIA (731.0/112.0)
|   cole_region_ubicacion = RIO MAYO: SOBRE MEDIA (829.0/333.0)
|   cole_region_ubicacion = SANQUIANGA: BAJO MEDIA (913.0/214.0)
|   cole_region_ubicacion = PIEDEMONTE COSTERO: BAJO MEDIA (244.0/92.0)
cole_naturaleza = NO OFICIAL
|   cole_region_ubicacion = PASTO: SOBRE MEDIA (789.0/206.0)

Number of Leaves  :         64

Size of the tree :      82
```

Para el caso de los estudiantes nariñenses, que están sobre la media en el desempeño en Matemáticas, el modelo tiene una precisión de predicción de 0.597, lo que quiere decir, que, del total de casos predichos que están sobre la media, el 59,7% están correctos. La sensibilidad (TPR) y exhaustividad (*Recall*) del modelo, es de 0.672, lo que indica, que el modelo clasifica correctamente al 67.2 % de los estudiantes que realmente están sobre la media. Por otra parte, la tasa de Falsos Positivos (*FP Rate*) del modelo es de 0.363, lo que significa, que el 36.3% de estudiantes que estaban bajo la media fueron clasificados como sobre la media. La medida F (*F-Measure* o *F1-Score*) es de 0.632 lo que significa, que la media armónica entre la precisión y la exhaustividad de los que están sobre la media no es tan buena, como la de los casos que están bajo la media, pero es mayor que 0.5.

El área bajo la curva ROC (*ROC Area*) del modelo es de 0.709 (mayor que 0.5), lo que indica, que el modelo tiene una alta probabilidad de que sea capaz de distinguir los valores de la clase sobre la media de los valores de la clase bajo la media, ya que es capaz de detectar más Verdaderos Positivos (TP) y Verdaderos Negativos (TN) que de Falsos Negativos (FN) y Falsos Positivos (FP).

En la combinación de estas métricas, se aprecia un mejor desempeño del modelo para los que están bajo la media.

Para escoger los patrones más representativos del árbol obtenido (Figura 15), se tuvo en cuenta, aquellos que superen un mínimo soporte de 0.5 % (i.e. mínimo número de ejemplos que llegan a un nodo) y una confianza mínima de 60% (i.e. el mínimo de ejemplos correctamente clasificados en cada nodo) con relación al total de ejemplos del conjunto de entrenamiento (16,843). Entre los patrones más importantes según su soporte están:

Patrón 1. Si el estudiante asiste a un colegio oficial y está ubicado en la Subregión Pacífico Sur de Nariño, entonces su desempeño académico

en la Competencia de Matemáticas de las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 16 % del total de estudiantes evaluados, se clasifican de esta manera. El 86.9 % de los estudiantes que tienen estas características, están correctamente clasificados y el 24.6 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 2. Si el estudiante asiste a un colegio oficial y está ubicado en la Subregión Sanquianga de Nariño, entonces su desempeño académico en la Competencia de Matemáticas de las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 5.4 % del total de estudiantes evaluados, se clasifican de esta manera. El 76.6 % de los estudiantes que tienen estas características, están correctamente clasificados y el 7.4 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 3. Si el estudiante asiste a un colegio oficial y está ubicado en la Subregión Rio Mayo de Nariño, entonces su desempeño académico en la Competencia de Matemáticas de las Pruebas Saber 5° tiene mayor probabilidad de estar sobre la media nacional. El 4.9 % del total de estudiantes evaluados, se clasifican de esta manera. El 60 % de los estudiantes que tienen estas características, están correctamente clasificados y el 6.7 % del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 4. Si el estudiante asiste a un colegio no oficial y está ubicado en la Subregión Pasto de Nariño, entonces su desempeño académico en la Competencia de Matemáticas de las Pruebas Saber 5° tiene mayor probabilidad de estar sobre la media nacional. El 4.7 % del total de estudiantes evaluados se clasifican de esta manera. El 73.9 % de los estudiantes que tienen estas características, están correctamente clasificados y el 7.9 % del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 5. Si el estudiante asiste a un colegio oficial y está ubicado en la Subregión Telembí de Nariño, entonces su desempeño académico en la Competencia de Matemáticas de las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 4.3 % del total de estudiantes evaluados se clasifican de esta manera. El 84.7 % de los estudiantes que tienen estas características, están correctamente clasificados y el 6.5 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 6. Si el estudiante asiste a un colegio oficial, está ubicado en la Subregión Pasto de Nariño y reprobó grados entonces su desempeño académico en la Competencia de Matemáticas de las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 4.1 % del total de estudiantes evaluados se clasifican de esta manera. El 61.9 % de los estudiantes que tienen estas características, están correctamente clasificados y el 4.5 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 7. Si el estudiante asiste a un colegio oficial y está ubicado en la Subregión de Piedemonte Costero de Nariño, entonces su desempeño académico en la Competencia de Matemáticas de las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 1.4 % del total de estudiantes evaluados se clasifican de esta manera. El 62.3 % de los estudiantes que tienen estas características, están correctamente clasificados y el 1.6 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 8. Si el estudiante asiste a un colegio oficial, está ubicado en la Subregión Los Abades de Nariño y en la zona rural, entonces su desempeño académico en la Competencia de Matemáticas de las Pruebas Saber 5° tiene mayor probabilidad de estar sobre la media nacional. El 1.4% del total de estudiantes evaluados, se clasifican de esta manera. El 61.8 %

de los estudiantes que tienen estas características, están correctamente clasificados y el 2 % del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 9. Si el estudiante asiste a un colegio oficial, está ubicado en la Subregión Juanambú de Nariño y reprobó grados, entonces su desempeño académico en la Competencia de Matemáticas de las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 1.4 % del total de estudiantes evaluados, se clasifican de esta manera. El 60 % de los estudiantes que tienen estas características, están correctamente clasificados y el 1.5 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 10. Si el estudiante asiste a un colegio oficial, está ubicado en la Subregión Centro de Nariño y pertenece al grupo etario de los 10 años, entonces su desempeño académico en la Competencia de Matemáticas de las Pruebas Saber 5° tiene mayor probabilidad de estar sobre la media nacional. El 1.2% del total de estudiantes evaluados se clasifican de esta manera. El 60 % de los estudiantes que tienen estas características, están correctamente clasificados y el 1.7 % del total de estudiantes observados que están sobre la media, cumplen este patrón.

Descubrimiento de Patrones de Desempeño General en las Pruebas Saber 5º de los Estudiantes del Departamento de Nariño

Para el descubrimiento de patrones de desempeño general en las Pruebas Saber 5º, presentada por los Estudiantes de Grado Quinto de Primaria del departamento de Nariño en el año 2017, se tomó el conjunto de datos *sbr5_narino_final* particionado en 70% para entrenamiento (16,843 registros) y 30% para prueba (7,219 registros). Se tomó como variable clase u objetivo, el atributo denominado *CLASE_DESEMP_GENERAL* (Tabla 43), con los valores sobre la media y bajo la media. En el conjunto de entrenamiento,

se observa 7,324 casos de estudiantes que están sobre la media y 9,519 casos de estudiantes que están bajo la media. Como parámetros de poda del árbol de decisión, se tomaron los valores para C=25% y para M=0.5%, según el proceso de ajuste de parámetros anteriormente descrito.

Con estos parámetros, se procedió a construir el modelo con el algoritmo J48. En la Figura 16, se muestra la Matriz de Confusión, con los parámetros de evaluación del modelo y en la Figura 17, se muestra el árbol de decisión obtenido, para el desempeño general en las Pruebas Saber 5º, una vez aplicado un proceso de pospoda según los parámetros de soporte mínimo y confianza mínima. Para este caso se tuvo en cuenta solo aquellos ramales del árbol que superen un mínimo soporte de 0.5 % (i.e. mínimo número de ejemplos que llegan a un nodo) y una confianza mínima de 60% (i.e. el mínimo de ejemplos correctamente clasificados en cada nodo).

Figura 16. Matriz de Confusión del árbol de desempeño general.

```
=== Summary ===

Correctly Classified Instances        4793               66.3942 %
Incorrectly Classified Instances      2426               33.6058 %
Kappa statistic                          0.3227
Mean absolute error                      0.4113
Root mean squared error                  0.4546
Relative absolute error                 83.5574 %
Root relative squared error             91.5768 %
Total Number of Instances             7219

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0,675    0,350    0,710     0,675    0,692      0,323    0,722     0,767     BAJO MEDIA
                 0,650    0,325    0,611     0,650    0,630      0,323    0,722     0,627     SOBRE MEDIA
Weighted Avg.    0,664    0,339    0,667     0,664    0,665      0,323    0,722     0,705

=== Confusion Matrix ===

   a    b   <-- classified as
 2728 1312 |    a = BAJO MEDIA
 1114 2065 |    b = SOBRE MEDIA
```

Analizando los resultados obtenidos en el modelo que se muestra en la Figura 16, sobre el desempeño general de los Estudiantes de Grado Quinto de las Instituciones Educativas de Básica Primaria de Nariño, que presentaron las Pruebas Saber 5º en el año 2017, se puede observar, que este clasifica correctamente al 66.4% del total de instancias probadas y clasifica incorrectamente al 33.6%.

Figura 17. Árbol podado para el desempeño general en las pruebas Saber 5º de Nariño

```
J48 pruned tree
------------------
cole_naturaleza = OFICIAL
|   cole_region_ubicacion = LA CORDILLERA
|   |   fami_educacionmadre = Bachillerato: BAJO MEDIA (211.0/84.0)
|   cole_region_ubicacion = PASTO
|   |   estu_reprobogrado = Si: BAJO MEDIA (693.0/262.0)
|   |   estu_reprobogrado = No
|   |   |   estu_edad = 11 anos: SOBRE MEDIA (1038.0/401.0)
|   cole_region_ubicacion = JUANAMBU
|   |   estu_reprobogrado = Si: BAJO MEDIA (240.0/95.0)
|   cole_region_ubicacion = PACIFICO SUR: BAJO MEDIA (2688.0/298.0)
|   cole_region_ubicacion = CENTRO
|   |   estu_genero = Mujer: SOBRE MEDIA (184.0/72.0)
|   cole_region_ubicacion = TELEMBI: BAJO MEDIA (731.0/91.0)
|   cole_region_ubicacion = RIO MAYO
|   |   cole_caracter = ACADEMICO: SOBRE MEDIA (469.0/168.0)
|   cole_region_ubicacion = SANQUIANGA: BAJO MEDIA (913.0/182.0)
|   cole_region_ubicacion = PIEDEMONTE COSTERO: BAJO MEDIA (244.0/93.0)
cole_naturaleza = NO OFICIAL
|   cole_region_ubicacion = OBANDO: SOBRE MEDIA (209.0/82.0)
|   cole_region_ubicacion = PASTO: SOBRE MEDIA (789.0/168.0)

Number of Leaves:  70
Size of the tree:      93
```

Evaluando el modelo con la Matriz de Confusión de la Figura 17 se observó que, de 7,219 casos de estudiantes probados, el 56 % (4,040 instancias) están bajo la media y 44 % (3,179 instancias) están sobre la media y el modelo predice al 53.2% (3,842 instancias) bajo la media y el 46.8% (3,377) sobre la media.

Por otra parte, el modelo predice correctamente a 2,728 casos de estudiantes nariñenses cuyo desempeño general están bajo la media (verdaderos negativos -TN) y a 2,065 casos de estudiantes nariñenses que están sobre la media (verdaderos positivos -TP). Además, 1,312 casos cuyo desempeño estaba bajo la media, el modelo los clasifica incorrectamente como sobre la media (falsos positivos-FP) y 1,114 casos cuyo desempeño estaba sobre la media, el modelo los clasifica incorrectamente como bajo la media (falsos negativos-FN). De acuerdo a esta matriz, el número de falsos positivos que el modelo clasifica es mayor que el número de falsos negativos, lo que significa que para este modelo es más costoso reconocer a aquellos estudiantes nariñenses cuyo desempeño general en las pruebas Saber 5º está bajo la media y que, para esta investigación, es el tipo de estudiantes que interesa.

En cuanto a las métricas de evaluación del modelo de clasificación, para el caso de los estudiantes nariñenses que están bajo la media en el desempeño general, el modelo tiene una precisión de predicción de 0.710, lo que quiere decir que, del total de casos predichos que están bajo la media, el 71 % están correctos. La sensibilidad (TPR) y exhaustividad (*Recall*) del modelo, es de 0.675, lo que indica que el modelo clasifica correctamente al 67.5 % de los estudiantes que realmente están bajo la media. Por otra parte, la tasa de Falsos Negativos (*FP Rate*) del modelo es de 0.350, lo que significa que el 35% de estudiantes que estaban sobre la media fueron clasificados como bajo la media. La medida F (*F-Measure* o *F1-Score*) es de 0.692 lo que significa que la media armónica entre la precisión y la exhaustividad de los que están bajo la media es buena, ya que es mayor que 0.5 y más cercano a 1.

Para el caso de los estudiantes nariñenses que están sobre la media en el desempeño general, el modelo tiene una precisión de predicción de 0.611, lo que quiere decir, que, del total de casos predichos que están sobre la media, el 61,1% están correctos. La sensibilidad (TPR) y exhaustividad

(*Recall*) del modelo, es de 0.650, lo que indica, que el modelo clasifica correctamente al 65 % de los estudiantes que realmente están sobre la media. Por otra parte, la tasa de Falsos Positivos (*FP Rate*) del modelo es de 0.325, lo que significa que el 32.5% de estudiantes que estaban bajo la media fueron clasificados como sobre la media. La medida F (*F-Measure* o *F1-Score*) es de 0.630 lo que significa que la media armónica entre la precisión y la exhaustividad de los que están sobre la media no es tan buena como la de los casos que están bajo la media, pero es mayor que 0.5.

El área bajo la curva ROC (*ROC Area*) del modelo es de 0.722 (mayor que 0.5), lo que indica, que el modelo tiene una alta probabilidad de distinguir los valores de la clase sobre la media de los valores de la clase bajo la media, ya que es capaz de detectar más Verdaderos Positivos (TP) y Verdaderos Negativos (TN) que de Falsos Negativos (FN) y Falsos Positivos (FP).

En la combinación de estas métricas, se aprecia un mejor desempeño del modelo para los que están bajo la media.

Para escoger los patrones más representativos del árbol obtenido (Figura 17), se tuvo en cuenta, aquellos que superen un mínimo soporte de 0.5 % (i.e. mínimo número de ejemplos que llegan a un nodo) y una confianza mínima de 60% (i.e. el mínimo de ejemplos correctamente clasificados en cada nodo) con relación al total de ejemplos del conjunto de entrenamiento (16,843). Entre los patrones más importantes según su soporte están:

Patrón 1. Si el estudiante asiste a un colegio oficial y está ubicado en la Subregión Pacifico Sur de Nariño, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 16 % del total de estudiantes evaluados se clasifican de esta manera. El 88.9 % de los estudiantes que tienen estas características, están correctamente clasificados y el 25.1 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 2. Si el estudiante asiste a un colegio oficial, está ubicado en la Subregión Pasto de Nariño, no reprobó grados y se encuentra en el grupo etario de los 11 años, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar sobre la media nacional. El 6.2 % del total de estudiantes evaluados, se clasifican de esta manera. El 61.4 % de los estudiantes que tienen estas características, están correctamente clasificados y el 8.7 % del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 3. Si el estudiante asiste a un colegio oficial y está ubicado en la Subregión Sanquianga de Nariño, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 5.4 % del total de estudiantes evaluados, se clasifican de esta manera. El 80.1 % de los estudiantes que tienen estas características, están correctamente clasificados y el 7.7 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 4. Si el estudiante asiste a un colegio no oficial y está ubicado en la Subregión de Pasto de Nariño, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar sobre la media nacional. El 4.7 % del total de estudiantes evaluados, se clasifican de esta manera. El 78.7 % de los estudiantes que tienen estas características, están correctamente clasificados y el 8.5 % del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 5. Si el estudiante asiste a un colegio oficial y está ubicado en la Subregión Telembí de Nariño, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 4.3 % del total de estudiantes evaluados, se clasifican de esta manera. El 87.6 % de los estudiantes que tienen estas características, están correctamente clasificados y el 6.7 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 6. Si el estudiante asiste a un colegio oficial, está ubicado en la Subregión Pasto de Nariño y reprobó grados, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 4.1 % del total de estudiantes evaluados, se clasifican de esta manera. El 62.2 % de los estudiantes que tienen estas características, están correctamente clasificados y el 4.5 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 7. Si el estudiante asiste a un colegio oficial, está ubicado en la Subregión Rio Mayo de Nariño y el carácter del colegio es académico, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar sobre la media nacional. El 2.8 % del total de estudiantes evaluados se clasifican de esta manera. El 64.2 % de los estudiantes que tienen estas características, están correctamente clasificados y el 4.1 % del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 8. Si el estudiante asiste a un colegio oficial y está ubicado en la Subregión de Piedemonte Costero de Nariño, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 1.4 % del total de estudiantes evaluados se clasifican de esta manera. El 61.9 % de los estudiantes que tienen estas características, están correctamente clasificados y el 1.6 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 9. Si el estudiante asiste a un colegio oficial, está ubicado en la Subregión Juanambú de Nariño y reprobó grados, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 1.4 % del total de estudiantes evaluados, se clasifican de esta manera. El 60.4 % de los estudiantes que tienen estas características, están correctamente clasificados y el 1.5 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 10. Si el estudiante asiste a un colegio oficial, está ubicado en la Subregión Cordillera de Nariño y el máximo nivel educativo de la madre es bachillerato, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar bajo la media nacional. El 1.3 % del total de estudiantes evaluados, se clasifican de esta manera. El 60.2 % de los estudiantes que tienen estas características, están correctamente clasificados y el 1.3 % del total de estudiantes observados que están bajo la media, cumplen este patrón.

Patrón 11. Si el estudiante asiste a un colegio no oficial y está ubicado en la Subregión de Obando de Nariño, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar sobre la media nacional. El 1.2 % del total de estudiantes evaluados, se clasifican de esta manera. El 60.8 % de los estudiantes que tienen estas características, están correctamente clasificados y el 1.7 % del total de estudiantes observados que están sobre la media, cumplen este patrón.

Patrón 12. Si el estudiante asiste a un colegio oficial, está ubicado en la Subregión Centro de Nariño y es de sexo femenino, entonces su Desempeño Académico General en las Pruebas Saber 5° tiene mayor probabilidad de estar sobre la media nacional. El 1.1 % del total de estudiantes evaluados, se clasifican de esta manera. El 60.9 % de los estudiantes que tienen estas características, están correctamente clasificados y el 1.5 % del total de estudiantes observados que están sobre la media, cumplen este patrón.

CAPÍTULO V
CONCLUSIONES Y TRABAJOS FUTUROS

El objetivo de este estudio fue descubrir factores asociados al desempeño académico de los Estudiantes de las Instituciones Educativas de Básica Primaria del departamento de Nariño quienes, cursando el Grado Quinto, presentaron las Pruebas Saber 5º.

Las fuentes de información fueron las bases de datos del Icfes, en las cuales se encuentra almacenada la información para el estudio, para lo cual, se seleccionaron los datos socioeconómicos, académicos e institucionales de los estudiantes nariñenses que presentaron las Pruebas Saber 5º en el año 2017; información que, para el momento del estudio, era la más actualizada.

El objetivo de la investigación consistió en el *Descubrimiento de Patrones de Desempeño Académico en las Pruebas Saber 5º*, para lo cual, se aplicó la metodología CRISP-DM y como técnica, la clasificación basada en árboles de decisión. Para el efeto, se evaluó el desempeño de diferentes algoritmos de árboles de decisión y se escogió el J48 por ser el que produjo mejores resultados.

Se realizó una caracterización socioeconómica e institucional de los estudiantes nariñenses que presentaron las Pruebas Saber 5º en el 2017; y se evaluó el desempeño académico en las Competencias de Lenguaje y Matemáticas, según la localidad, a través de un ranking, considerando el municipio y la subregión del departamento de Nariño.

Se consideró que el buen desempeño académico en las Pruebas Saber 5°, corresponde a los puntajes por encima de la media nacional, y un bajo desempeño, a los puntajes por debajo de dicha. Se encontró que, es mayor el porcentaje de estudiantes del departamento de Nariño que

tienen un desempeño académico bajo, en comparación con el porcentaje de estudiantes con un buen desempeño; hecho que se presenta tanto en Lenguaje como en Matemáticas y también en el puntaje general.

En el departamento de Nariño, los factores determinantes asociados al buen desempeño académico de los estudiantes en las Pruebas Saber 5º, corresponden a estudiantes de un colegio oficial, situado en la Subregión de Pasto, no haber reprobado grados y pertenecer al grupo etario de los 11 años. Para el caso de que el estudiante pertenezca a una institución privada, entonces, las Subregiones de Pasto y Obando son otro factor a tener en cuenta para el buen desempeño académico.

Por otra parte, los factores determinantes asociados al bajo desempeño académico de los estudiantes en las Pruebas Saber 5º del departamento de Nariño, son los de estudiar en un colegio oficial y estar situado en la Subregión Pacífico Sur.

Otros factores asociados en menor grado a este desempeño son los de pertenecer a las Subregiones de Telembí, Sanquianga y Piedemonte Costero; y en general, pertenecer a la zona Pacífica del departamento de Nariño.

Finalmente, el impacto esperado de esta investigación podría ser significativo en varios aspectos, tanto a nivel educativo como en términos de políticas públicas y gestión de recursos.

A continuación, se presentan los principales impactos esperados:

1. **Mejoramiento de la Calidad Educativa**:

- Los hallazgos sobre los factores que favorecen un buen desempeño académico en las subregiones de Pasto y Obando podrían orientar las intervenciones en otras zonas del departamento, especialmente

en las áreas con bajo desempeño, como la Subregión Pacífico Sur y otras áreas del Pacífico nariñense.

- Al identificar que estudiantes de 11 años, en ciertas subregiones y en instituciones privadas, tienden a tener un mejor rendimiento, las instituciones educativas y autoridades locales podrían desarrollar programas específicos para replicar prácticas exitosas en las zonas con bajo desempeño.

2. **Diseño de políticas y asignación de recursos**:

- Las autoridades locales y departamentales podrían enfocar recursos y programas de apoyo en zonas con mayor porcentaje de bajo rendimiento, como las subregiones de Telembí, Sanquianga y Piedemonte Costero. Esto podría implicar, por ejemplo, la mejora de infraestructura, la capacitación de docentes y el aumento de recursos en las escuelas de estas regiones.

- Se podría establecer una estrategia para reducir las tasas de reprobación en la educación primaria, considerando que uno de los factores de buen rendimiento es no haber reprobado grados. Programas de apoyo y refuerzo académico desde los primeros grados podrían impactar positivamente en el desempeño en pruebas futuras.

3. **Aumento en la equidad educativa**:

- El estudio subraya desigualdades entre subregiones y tipos de institución, por lo que sus conclusiones podrían servir como base para intervenciones específicas que busquen reducir las brechas de desempeño en el departamento de Nariño.

- Al comprender que los estudiantes en colegios oficiales de ciertas subregiones muestran un rendimiento inferior, las políticas públicas pueden enfocarse en brindar apoyo adicional a estas instituciones para nivelar el acceso a una educación de calidad y mejorar los resultados.

4. **Optimización del seguimiento del desempeño académico**:

- Recomendación de futuros estudios que analicen los factores de desempeño en las regiones de bajo desempeño, como la Subregión Pacífico Sur y las áreas de Telembí, Sanquianga y Piedemonte Costero, donde el estudio detectó mayores dificultades académicas. Estos hallazgos permiten que los recursos públicos se distribuyan de manera más eficiente y dirigida, maximizando el impacto de las inversiones en educación.

5. **Intervenciones para la reducción de brechas educativas**:

- La diferencia en el desempeño académico entre subregiones sugiere la existencia de brechas en la calidad educativa en el departamento de Nariño. Los resultados de la investigación pueden orientar programas para reducir estas brechas, como el fortalecimiento de la infraestructura educativa, la capacitación de docentes y la implementación de programas de apoyo pedagógico específico en las zonas menos favorecidas.

- El estudio resalta cómo factores como el tipo de institución (oficial o privada) y la ubicación geográfica inciden en el desempeño académico, lo cual podría justificar la implementación de programas de equidad para mejorar las oportunidades educativas en sectores desfavorecidos.

6. **Promoción de la continuidad y calidad educativa en secundaria y media**:

- Al sugerirse como trabajo futuro un análisis de los factores asociados en las pruebas Saber 9º y Saber 11º, se podría evaluar el valor agregado de la educación secundaria y media. Esto sería clave para garantizar que los estudiantes no solo mejoren su rendimiento en los grados superiores, sino también para analizar la efectividad de los programas de refuerzo en las áreas de menor rendimiento.

- La investigación también podría motivar la creación de estrategias de seguimiento para estudiantes desde la primaria hasta la media, identificando cómo los factores asociados a un buen rendimiento en primaria pueden mantenerse y fortalecerse en los grados superiores.

7. **Fortalecimiento de la planeación curricular y pedagógica**:

- La información sobre factores de éxito en algunas subregiones puede ayudar a diseñar modelos pedagógicos adaptados al contexto local que respondan a las características y necesidades de cada subregión, especialmente en áreas con resultados académicos más bajos.

- Los resultados también podrían influir en la planificación de currículos y estrategias pedagógicas en primaria, de manera que promuevan una mejor preparación para los grados superiores, con énfasis en las competencias de lenguaje y matemáticas.

Esta investigación no solo ofrece un diagnóstico valioso del desempeño académico en Nariño, sino que también sienta las bases para mejoras

estratégicas en el sistema educativo departamental. La identificación de factores asociados al rendimiento académico en primaria permitirá a las autoridades diseñar intervenciones más efectivas, reducir las brechas educativas y asegurar que los estudiantes en todo el departamento tengan mayores oportunidades de éxito en su trayectoria educativa.

Se podría afirmar que la aplicación de la metodología CRISP-DM y el uso de árboles de decisión abren camino para que las instituciones de educación y los gobiernos locales utilicen herramientas de minería de datos en futuras investigaciones. Esto permitiría analizar patrones y tendencias de desempeño, facilitando una toma de decisiones informada y basada en datos.

Finalmente, con base en los resultados de la investigación, que evidencian un desempeño desigual entre las subregiones del Departamento, se podrían implementar programas específicos para reducir las brechas educativas, priorizando las zonas de bajo desempeño, como la Subregión Pacífico Sur, Telembí, Sanquianga y Piedemonte Costero. Estas intervenciones podrían incluir la mejora de infraestructura educativa, la capacitación docente orientada a competencias específicas y la provisión de recursos pedagógicos innovadores. Además, se podría promover la equidad educativa mediante políticas que garanticen un acceso igualitario a una educación de calidad en las instituciones oficiales, fortaleciendo las capacidades de los estudiantes para superar las barreras sociales y geográficas que afectan su desempeño.

Por otra parte, para mejorar el rendimiento en competencias de Lenguaje y Matemáticas, se pueden desarrollar currículos adaptados al contexto socioeconómico de cada subregión. Estos currículos podrían incorporar metodologías pedagógicas innovadoras que promuevan el aprendizaje activo y autónomo. Además, se podría implementar un sistema de seguimiento académico integral, desde la primaria hasta la media, permitiendo

monitorear el impacto de las intervenciones y asegurar la continuidad del progreso educativo. En el marco de esta política, también se podrían utilizar herramientas avanzadas, como la minería de datos, para identificar patrones de desempeño, optimizando la asignación de recursos y la formulación de estrategias basadas en evidencia.

Se plantea como trabajos futuros, complementar este estudio, utilizando técnicas descriptivas de minería de datos, que permitan relacionar, cuales atributos se presentan juntos, asociados al desempeño académico en las Pruebas Saber 5° y cómo se agrupan los individuos de acuerdo a su rendimiento en dichas pruebas.

Además, sería recomendable, realizar estudios sobre la relación entre los factores asociados al desempeño académico de los estudiantes en las Pruebas Saber 5°, Saber 9º y Saber 11º, con el fin de determinar el valor agregado de la Educación Básica Secundaria y media en la formación de los bachilleres colombianos y específicamente del departamento de Nariño.

REFERENCIAS

Agasisti, T. y Cordero, J. M. (2014). The determinants of repetition rates in European countries: insights from an empirical analysis using PIRLS 2011 data. Instituto Nacional de Evaluación Educativa - INEE.

Ahumada, V., Zapata, M. y Varela, R. (2014). Aprendiendo a investigar. Investigando el carnaval de Barranquilla. [Trabajo de maestría, Universidad del Atlántico]. https://repositorio.unicartagena.edu.co/handle/11227/1696

Arum, R. y Velez, M. (Eds.). (2012). Improving learning environments: School discipline and student achievement in comparative perspective. Stanford University Press.

Ayala, J. (2015). Evaluación externa y calidad de la educación en Colombia. Documentos de trabajo sobre economía regional. Banco de la República.

Azevedo, A. y Santos, M. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. IADIS. https://www.researchgate.net/publication/220969845_KDD_semma_and_CRISP-DM_A_parallel_overview

Betts, J. R. y Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. Economics of Education Review, 19(1), 1-15.

Bourdieu, P. y Passeron, J. (2001). La reproducción. Elementos para una teoría del sistema de enseñanza. Editorial Popular.

Cárcamo, C. y Mola, J. (2012). Diferencias por sexo en el desempeño académico en Colombia: Un análisis regional. Revista Economía & Región, 6(1). 133-169.

Celis, M. T., Jiménez, Ó. A., y Jaramillo, J. F. (2012). ¿Cuál es la brecha de la calidad educativa en Colombia en la educación media y en la superior? En Icfes. (Ed.).

Estudios sobre calidad de la educación en Colombia (pp. 67-98). Instituto Colombiano para la Evaluación de la Educación Icfes.

Chapman, P., Clinton, J., Kerber, R, Khabaza, T., Reinartz, T., Shearer, C., y Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. CRISP-DM consortium.

Chaux, E. (2009). Citizenship Competencies in the Midst of a Violent Political Conflict: The Colombian Educational Response. Harvard Educational Review, 79(1), 84–93. Disponible en: http://doi.org/10.17763/haer.79.1.d2566q027573h219

Cohen, J. (1998). Statistical Power Analysis for the Behaviorial Sciences. Second Edition. New Lawrence Erlbaum Associates, Publishers.

Coleman, J. S. (1966). Equality of educational opportunity. Dept. of Health, Education, and Welfare, Office of Education.

Congreso de la República de Colombia (1991). Código de Procedimiento Civil [C.P.C]. Decreto 1400 de 1970. Colombia. http://www.secretariasenado.gov.co/senado/basedoc/codigo_procedimiento_civil.html

Cunha, F. & Heckman, J. (2007). The technology of skill formation (No. w12840). National Bureau of Economic Research.

Frawley, W.J., Piatetsky-Shapiro, G. & Matheus, C.J. (1992). Knowledge discovery in databases: An overview. AI Magazine , 13 (3), 57. https://doi.org/10.1609/aimag.v13i3.1011

FEDESARROLLO. Centro de investigación económica y social. (2014). La educación básica y media en Colombia: retos en equidad y calidad. FEDESARROLLO. https://www.repository.fedesarrollo.org.co/handle/11445/190

Gallardo, J. (2009). Metodología para la definición de requisitos en proyectos de data mining. [Tesis Doctoral, Facultad de Informática (UPM)]. https://doi.org/10.20868/UPM.thesis.1946

Gómez, L. M. y Macedo, J. C. (2010). Importancia de las TIC en la Educación. Investigación Educativa, 14(25), 209–226. http://revistasinvestigacion.unmsm.edu.pe/index.php/ educa/article/view/4776

Green, J., Liem, G. A. D., Martin, A. J., Colmar, S., Marsh, H. W., & McInerney, D. (2012). Academic motivation, self-concept, engagement, and performance in high school: Key processes from a longitudinal perspective. Journal of Adolescence, 35(5), 1111-1122. http://dx.doi.org/10.1016/j.adolescence.2012.02.016

Gunnarsson, V., Orazem, P. F. & Sánchez, M. A. (2006). Child Labor and School Achievement in Latin America. The World Bank Economic Review, 20(1), 31-54. https://elibrary.worldbank.org/doi/10.1093/wber/lhj003

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. En M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), Handbook of reading research, Vol. 3 (pp. 403–422). Lawrence Erlbaum Associates Publishers.

Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional contexts for engagement and achievement in reading. In Christenson, S., Reschly, A., Wylie, C. (eds.), Handbook of Research on Student Engagement (pp 601–634). Springer.

Gutiérrez, Y. (2014). Relación entre la estructura familiar y el rendimiento académico en el área de matemáticas, en estudiantes de tercero y quinto de primaria. Instituto Latinoamericano de altos estudios. https://libroselectronicos.ilae.edu.co/index.php/ilae/catalog/book/43

Hall, M., Frank, E., & Witten. (2011). Practical Machine Learning Tools and Techniques. Morgan Kaufmann.

Han, J. & Kamber, M. (2001). Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers.

Han, J., Kamber, M. & Pei, J. (2011). Data Mining: Concepts and Techniques (3 edition). Morgan Kaufmann Publishers. https://archive.org/details/the-morgan-kaufmann-series-in-data-management-systems-jiawei-han-micheline-kambe

Heckman, J. J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. Science, 312(5782), 1900-1902. https://www.doi.org/10.1126/science.1128898

Hernández, E. & Lorente, R. (2009). Minería de datos aplicada a la detección de Cáncer de Mama. Universidad Carlos III de Madrid. https://www.researchgate.net/publication/265891193_Minera_de_datos_aplicada_a_la_deteccion_de_Cancer_de_Mama

Hernández, J., Ramírez, M. & Ferri, C. (2005). Introducción a la Minería de Datos. Pearson Prentice Hall.

Hoffer, T. B. (1992). Middle school ability grouping and student achievement in science and mathematics. Educational Evaluation and Policy Analysis, 14(3), 205-227. https://doi.org/10.3102/01623737014003205

Holmes, C. (1989). Grade Level Retention Effects: A Meta Analysis of Research Studies in L.A. En L. Shepard & M. Smith. (Eds.), Flunking Grades: Research and Policies on Retention (pp. 16-33). The Falmer Press.

Icfes. (2009). Lineamientos generales Saber 5° y 9°. Instituto Colombiano para la Evaluación de la Educación.

Icfes. (2009b). Informe técnico Saber 5o. y 9o. 2009. Instituto Colombiano para la Evaluación de la Educación.

Icfes. (2011). Informe técnico Saber 5° y 9°. Instituto Colombiano para la Evaluación de la Educación.

Icfes. (2014a). Pruebas Saber 3°, 5° y 9°. Lineamientos para las aplicaciones muestral y censal 2014. Instituto Colombiano para la Evaluación de la Educación. https://www.atlantico.gov.co/images/stories/adjuntos/educacion/lineamientos_muestral_censal_Saber359_2014.pdf

Icfes. (2014b). Alineación del examen Saber 11°. Lineamientos generales 2014 - 2 Sistema Nacional de Evaluación Estandarizada de la Educación. Instituto Colombiano para la Evaluación de la Educación.

Icfes. (2016a). Marco de factores asociados Saber 3°, 5° y 9°. Instituto Colombiano para la Evaluación de la Educación.

Icfes. (2016b). Guías de divulgación. Instituto Colombiano para la Evaluación de la Educación. https://web.archive.org/web/20220302133453/https://www.icfes.gov.co/web/guest/divulgacion-Saber-359-2016

Icfes. (2017). Experiencia acumulada en el levantamiento de datos de factores asociados 2012-2015, Saber 3°, 5° y 9°. Instituto Colombiano para la Evaluación de la Educación.

Icfes. (2022). Guía de orientación grado 5º Competencias comunicativas en Lenguaje: Lectura. Instituto Colombiano para la Evaluación de la Educación.

Icfes. (2022b). Prueba nacional muestral y controlada 2022 . Saber 9º. °. Instituto Colombiano para la Evaluación de la Educación.

Icfes. (s.f.). Documentación de la prueba Saber 3°, 5° y 9°. Instituto Colombiano para la Evaluación de la Educación. https://www.icfes.gov.co/documents/39286/2386763/Documentacion+Saber+359.pdf

Jenkins, E. J. & Bell, C. C. (1994). Violence among inner city high school students and post traumatic stress disorder. En S. Friedman. (Ed.), Anxiety disorders in African Americans (pp. 76-88). Springer.

Kerckhoff, A. C. (1986). Effects of ability grouping in British secondary schools. American Sociological Review, 51, 842–858.

Larose, D. & Larose, Ch. (2014). Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons, Inc Publishers. https://www.doi.org/10.1002/9781118874059

Ley 115 de 1994. (febrero 8 de 1994). Por la cual se expide la ley general de educación. Congreso de Colombia. http://secretariasenado.gov.co/senado/basedoc/ley_0115_1994.html

Marmolejo, G., Tarapuez, L. & Blanco, H. (2019). Geometría y Medición en las Pruebas Saber-Grado Quinto ¿Qué evalúan? Revista EIA, 16(32), 55-64.

Martinez, A. (2019). Subregiones departamento de Nariño. Toda Colombia. https://www.todacolombia.com/departamentos-de-colombia/narino/subregiones.html

Martínez, S. (2015). Pruebas Saber de lenguaje 3° y 5°: Posibilidades y retos desde la perspectiva de la evaluación formativa. Universidad Pedagógica Nacional.

MEN. (1998). Matemáticas: Lineamientos curriculares. Editorial Panamericana Formas e impresos.

MEN. (2006). Estándares Básicos de Competencias en Lenguaje, Matemáticas, Ciencias y Ciudadanas: Guía sobre lo que los estudiantes deben Saber y Saber hacer con lo que aprenden. Ministerio de Educación Nacional.

MEN. (2009). Decreto 1290 de 2009. Por el cual se reglamenta la evaluación del aprendizaje y promoción de los estudiantes de los niveles de educación básica y media. Ministerio de Educación Nacional. https://www.mineducacion.gov.co/1621/articles-187765_archivo_pdf_decreto_1290.pdf

MEN. (2014). Informe de Gestión y Rendición de Cuentas Vigencia 2014. Ministerio de Educación Nacional. https://www.mineducacion.gov.co/1621/articles-355956_Informe_gestion_2014_completo.pdf

MEN. (2015). Pruebas Saber 3°, 5° y 9°. Lineamientos para las aplicaciones muestral y Censal 2015. Ministerio de Educación Nacional.

MEN. (2016). Saber 5º Lineamientos para las aplicaciones muestral y censal 2016. Ministerio de Educación Nacional.

MEN. (2017). Guía de orientación Saber 5°. Ministerio de Educación Nacional. https://www.studocu.com/es-ar/document/universidad-nacional-de-catamarca/matematica/guia-de-orientacion-Saber-5-2017/5084470

Moine, J. M., Haedo, A. & Gordillo, S. (2011). Estudio comparativo de metodologías para minería de datos. Red de Universidades con Carreras en Informática. http://sedici.unlp.edu.ar/handle/10915/20034

Moiné, J. & Haedo, A. (2015). Una herramienta para la evaluación y comparación de metodologías de minería de datos. Universidad Nacional del Noroeste de la Provincia de Buenos Aires. https://sedici.unlp.edu.ar/bitstream/handle/10915/50428/Documento_completo.pdf?sequence=1

Molina, J. & García, J. (2006). Técnicas de Análisis de Datos: aplicaciones prácticas utilizando Mocrosoft Excel y Weka. Universidad Carlos III de Madrid. https://matema.ujaen.es/jnavas/web_recursos/archivos/weka%20master%20recursos%20naturales/apuntesAD.pdf

Murnane, R. & Ganimian, A. (2014). Improving Education in Developing Countries: Lessons from Rigorous Impact Evaluations. Review of Educational Research, 86(3), 719-755.

Naranjo, A. (2011). Evaluación integral para una educación de calidad. Revista Internacional Magisterio. Educación y Pedagogía, 9(51), 48-54.

OECD. (2003). Learners for life. Student approaches to learning. Results from PISA 200. OECD Publishing. https://doi.org/10.1787/9789264103917-en

OECD. (2005). School Factors Related to Quality and Equity. Results from PISA 2000. OECD Publishing. https://doi.org/10.1787/9789264008199-en

OECD. (2010a). PISA 2009 Results: Learning to Learn – Student Engagement, Strategies and Practices (Volume III). OECD Publishing. http://dx.doi.org/10.1787/9789264083943-en

OECD. (2010b). PISA 2009 Results: What Makes a School Successful? – Resources, Policies and Practices (Volume IV). OECD Publishing. http://dx.doi.org/10.1787/9789264091559-en

OECD. (2011a). Education at a Glance 2011: OECD Indicators. OECD Publishing. http://dx.doi.org/10.1787/eag-2011-en

OECD. (2011b). PISA Against the Odds Disadvantaged Students Who Succeed in School. OECD Publishing. https://doi.org/10.1787/9789264090873-en

OECD. (2016a). Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy. OECD Publishing. https://doi.org/10.1787/9789264255425-en

OECD. (2016b). Education in Colombia, Reviews of National Policies for Education. OECD Publishing. https://doi.org/10.1787/9789264250604-en

Patall, E. A., Cooper, H. & Allen, A. B. (2010). Extending the School Day or School Year: A Systematic Review of Research (1985–2009). Review of Educational Research, 80(3), 401-436. doi:10.3102/0034654310377086.

Posada, J. & Mendoza, F. (2014). Determinantes del logro académico de los estudiantes de grado 11 en el periodo 2008-2010. Una perspectiva de género y región. Icfes, Ministerio de Educación Nacional.

Pyle, D. (2003). Business Modeling and Data Mining. Morgan Kaufmann Publishers. https://doi.org/10.1016/B978-1-55860-653-1.X5000-3

Quinlan, J. (1993). Programs for machine learning. Morgan Kaufmann Editors.

Reschly, A. L. & Christenson, S. L. (2013). Grade retention: Historical perspectives and new research. Journal of school psychology, 51(3), 319-322.

Roderick, M. (1994). Grade Retention and School Dropout: Investigating the Association. American Educational Research Journal, 31(4), 729-759.

Rodríguez, D. & Ahumada, V. (2017). Análisis de los resultados de las pruebas externas Saber en matemáticas de un grupo de estudiantes de quinto grado. Revista Paradigma, 38(2), 275 – 290.

Ryan, R. M. & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. University of Rochester.

Sattler, K. & Dunemann, O. (2001). SQL Database Primitives for Decision Tree Classifiers. En H. Paques, L. Liu, & D. Grossman. (Eds.), The 10th ACM International Conference on Information and Knowledge Management (pp. 379-386). ACM.

Scheerens, J. (2000). Improving school effectiveness. Fundamentals of Educational Planning, United Nations Educational, Scientific and Cultural Organization.

Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. The Journal of Data Warehousing, 5(4), 13-22.

Sirin, S. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. Review of educational research, 75(3), 417-453. https://doi.org/10.3102/00346543075003417

Spicker, P., Leguizamón, S. & Gordon, D. (2009). Pobreza: Un glosario internacional. Primera edición. Consejo Latinoamericano de Ciencias Sociales. https://www.crop.org/publications/clacso/pobreza-un-glosario-internacional.aspx

Timarán, R. & Millán, M. (2006). New algebraic operators and SQL primitives for mining classification rules. [conferencia]. Computational Intelligence, San Francisco, California. http://www.actapress.com/PaperInfo.aspx?PaperID=29048&reason=500

Timarán, R., Jiménez, J. & Calderón, A. (2017). Detección de patrones de deserción estudiantil con minería de datos. Editorial Universidad de Nariño.

Timarán, R., Caicedo, J. & Hidalgo, A. (2019). Identificación de Factores Asociados al Desempeño Académico en Matemáticas en las Pruebas Saber 11° Aplicando Minería de Datos Educativa. LACCEI, Inc.

Timarán, R., Caicedo, J. & Hidalgo, A. (2021). Minería de datos educativa para el descubrimiento de factores asociados al desempeño académico en las Pruebas Saber 11º. Editorial Universidad de Nariño.

Torres, J. Pachajoa, L. & Pantoja, R. (2014). Resultados de las Pruebas Saber en el grado quinto del área de las ciencias naturales en tres instituciones educativas oficiales del municipio de Pasto. Revista Fedumar Pedagogía y Educación, 1(1), 55-69, 2014.

Trautwein, U. & Möller, J. (2016). Self-Concept: Determinants and Consequences of Academic Self-Concept in School Contexts. En A. Lipnevich, F. Preckel, & R. Roberts. (Eds.), The Springer Series on Human Exceptionality (pp. 187–214). Springer. https://doi.org/10.1007/978-3-319-28606-8_8

Unesco. (2010). Factores asociados al logro cognitivo de los estudiantes en América Latina y el Caribe. Oficina Regional de Educación de la Unesco para América Latina y el Caribe.

Unesco. (2015). Factores Asociados. Informe de resultados Tercer Estudio Regional Comparativo y Explicativo. Oficina Regional de Educación para América Latina y el Caribe.

Vasco, C.E. (2003). Objetivos específicos, indicadores de logros y competencias ¿y ahora estándares? Educación y Cultura, 62, 33-41.

Weinstein, C. E., Husman, J., & Dierking, D. R. (2000). Self-regulation interventions with a focuson learning strategies. En M. Boekaerts, P. Pintrich, & M. Zeidner. (Eds.), Andbook of selfregulation (pp. 727–747). Academic Press.

Witten, I., Frank, E., & Hall, M., (2011). Data Mining: Practical Machine Learning Tools and Techniques, Third Edition. Morgan Kaufmann. https://doi.org/10.1016/C2009-0-19715-5

Lista de figuras

Figura 1. Modelo Ecológico 45

Figura 2. Modelo de Eficacia Escolar de Scheerens 46

Figura 3. Modelo CIPP 48

Figura 4. Modelo de Factores Asociados adoptado por el Icfes. 49

Figura 5. Resultados del nivel de desempeño de los estudiantes nariñenses en la competencia de Lenguaje de las Pruebas Saber 5° 78

Figura 6. Resultados del nivel de desempeño de los estudiantes nariñenses en las competencia de Matemáticas de las Pruebas Saber 5° 78

Figura 7. Clasificación de los municipios de Nariño de acuerdo con los resultados en Lenguaje de las Pruebas Saber 5° 80

Figura 8. Clasificación de las subregiones de Nariño de acuerdo con los resultados en Lenguaje de las Pruebas Saber 5° 81

Figura 9. Clasificación de los municipios de Nariño de acuerdo con los resultados en matemáticas de las Pruebas Saber 5° 82

Figura 10. Clasificación de las subregiones de Nariño de acuerdo con los resultados en Matemáticas de las Pruebas Saber 5° 83

Figura 11. Fases de la Metodología CRISP-DM. 104

Figura 12. Matriz de Confusión del árbol para la competencia de Lenguaje 122

Figura 13. Árbol podado para la competencia de Lenguaje del Saber 5º de Nariño. 123

Figura 14. Matriz de confusión para el árbol de la competencia de Matemáticas. 131

Figura 15. Árbol podado para la competencia de Matemáticas del Saber 5º de Nariño. 133

Figura 16. Matriz de Confusión del árbol de desempeño general. 138

Figura 17. Árbol podado para el desempeño general en las pruebas Saber 5º de Nariño 139

Lista de tablas

Tabla 1. Competencias evaluadas en cada Prueba Saber 17

Tabla 2. Antecedentes de las Pruebas Saber 3°, 5° y 9°: Pruebas Muestrales de Lenguaje, Matemáticas y Ciencias 21

Tabla 3. Antecedentes de las Pruebas Saber 3°, 5° y 9°: Pruebas Censales de Lenguaje, Matemáticas y Ciencias Naturales, Ciencias Sociales y Competencias Ciudadanas 21

Tabla 4. Fases de la Evaluación 22

Tabla 5. Competencia Comunicativa Lectora – Grados 4° y 5° 26

Tabla 6. Competencia Comunicativa Escritora – Grados 4° y 5° 27

Tabla 7. Competencia: Comunicación, Representación y Modelación - 4° y 5° 31

Tabla 8. Competencia: Razonamiento y Argumentación - 4° y 5° Grados 32

Tabla 9. Competencia: Planteamiento y Resolución de Problemas - 4° y 5° Grados 33

Tabla 10. Competencia: Uso del Conocimiento Científico - 1° a 5° Grados 37

Tabla 11. Competencia: Explicación de Fenómenos - 1° a 5° Grados 38

Tabla 12. Competencia: Indagación - 1° a 5° Grados 39

Tabla 13. Competencia: Pensamiento Ciudadano 41

Tabla 14. Competencias Evaluadas y Aspectos Destacados - Grado 5° 42

Tabla 15. Histórico Dimensión Contexto: Características de los Estudiantes y sus Familias 51

Tabla 16. Histórico Dimensión Contexto: Características de las Instituciones Educativas 52

Tabla 17. Histórico Dimensión Insumos 53

Tabla 18. Histórico Dimensión Procesos 54

Tabla 19. Histórica Dimensión Resultados Educativos 54

Tabla 20. Características Socioeconómicas de los Estudiantes Nariñenses que Presentaron las Pruebas Saber 5° 72

Tabla 21. Características Institucionales de los Estudiantes Nariñenses que Presentaron las Pruebas Saber 5° *74*

Tabla 22. Efecto de las Variables Socioeconómicas en el Desempeño de Lenguaje en las Pruebas Saber 5° del Departamento de Nariño..... *85*

Tabla 23. Efecto de las Variables Socioeconómicas en el Desempeño de Matemáticas en las Pruebas Saber 5° del Departamento de Nariño..... *87*

Tabla 24. Efecto de las Variables Institucionales en el Desempeño de Lenguaje en las Pruebas Saber 5° del Departamento de Nariño..... *89*

Tabla 25. Efecto de las Variables Institucionales en el Desempeño de Matemáticas en las Pruebas Saber 5° del Departamento de Nariño..... *91*

Tabla 26. Diccionario de Datos del Conjunto de Datos sbr5_narino. *93*

Tabla 27. Diccionario de Datos de los Conjuntos de Datos sbr5_narino_final..... *108*

Tabla 28. Valores Discretizados del Atributo estu_edad..... *110*

Tabla 29. Subregiones de Nariño..... *111*

Tabla 30. Cálculo del Índice para fami_electrodomesticos *112*

Tabla 31. Cálculo del Índice para fami_tic *113*

Tabla 32. Evaluación Algoritmos de Árboles de Decisión..... *117*

Tabla 33. Ajuste de Parámetros de Poda para sbr5_narino_final..... *119*

Tabla 34. Particionamiento del Conjunto de Datos sbr5_narino_final *120*

Tabla 35. Distribución por Clases de sbr5_narino_final *121*

ACERCA DE LOS AUTORES

SILVIO RICARDO TIMARÁN PEREIRA

Doctor en Ingeniería con énfasis en Ciencias de la Computación. Master of Science en Ingeniería, Especialista en Multimedia Educativa e Ingeniero de Sistemas y Computación. Profesor Titular adscrito al departamento de Sistemas de la facultad de Ingeniería de la Universidad de Nariño. Director del Grupo de Investigación Aplicada en Sistemas -GRIAS y Coordinador de la Maestría en Gestión de Tecnologías de la Información y del Conocimiento MGTIC.

ARSENIO HIDALGO TROYA

Magister en Ciencias Estadísticas, Especialista en Proyectos de Desarrollo, Economista y Licenciado en Matemáticas y Física. Profesor Titular adscrito al departamento de Matemáticas y Estadística de la facultad de Ciencias Naturales y Exactas de la Universidad de Nariño. Director del Centro en Estudios en Salud CESUN.

SEGUNDO JAVIER CAICEDO ZAMBRANO

Doctor en Ciencias de la Educación, Magister en Software Libre, Especialista en Computación para la Docencia, Especialista en Multimedia Educativa y Licenciado en Matemáticas y Física. Profesor Asociado adscrito al departamento de Matemáticas y Estadística de la facultad de Ciencias Naturales y Exactas de la Universidad de Nariño.

êditorial
Universidad de **Nariño**

Año de publicación 2025
San Juan de Pasto - Nariño - Colombia

En este libro se presentan los resultados del proyecto de investigación denominado "Aplicación de técnicas de minería de datos para el descubrimiento de factores asociados al desempeño académico en las Pruebas Saber 5° de los estudiantes de las Instituciones Educativas del Departamento de Nariño". El objetivo de este estudio fue descubrir factores asociados al desempeño académico de los estudiantes de las Instituciones Educativas de Básica Primaria del Departamento de Nariño, que cursando el grado Quinto presentaron las pruebas Saber 5° en el año 2017. El estudio se realizó aplicando técnicas de minería de datos, a un repositorio, integrado por los datos socioeconómicos, académicos e institucionales, extraídos de las bases de datos del ICFES. Se utilizó la metodología CRISP-DM, guía ampliamente empleada en el desarrollo de proyectos de minería de datos. Se construyó, limpió y transformó un repositorio de datos para la minería de datos. Se construyó un modelo de clasificación basado en árboles de decisión y se descubrieron patrones asociados al buen o mal desempeño académico de los estudiantes, tanto a nivel nacional como del Departamento de Nariño, en las competencias de Lenguaje y Matemáticas, así como al desempeño general, que evalúa en Saber 5° en el año 2017.

El conocimiento descubierto, se incorporará al existente y se podrá integrar a los procesos de toma de decisiones de los organismos gubernamentales e instituciones educativas de Colombia y del Departamento de Nariño, así como del MEN, en lo relacionado con la implementación de estrategias y planes en sus respectivas campos de actuación, para mejorar la calidad de la Educación Básica Primaria en Colombia y específicamente en el Departamento de Nariño.

ISBN: 978-628-7771-66-6
9 786287 771666

Universidad de Nariño
FUNDADA EN 1904

Universidad de Nariño
ACREDITADA EN ALTA CALIDAD
RESOLUCIÓN MEN 000022 - ENERO 11 DE 2023

120 Años
Universidad de Nariño

êditorial
Universidad de Nariño