Declive en el rendimiento en comprensión de lectura en cuarto grado escolar: un fenómeno presente

Autor: Hugo Andres Sanchez Sanchez

Tesis para optar por el título de Magíster en Psicología con énfasis en evaluación y mejora de procesos educativos

Directora: Olga Rosalba Rodríguez Jiménez

Universidad Nacional de Colombia

2022

Viene a mi memoria una clase en la Universidad que tanto amo, la Universidad Nacional de Colombia, en ella, una docente nos reunió y nos dijo que cada uno mostrara una frase, una que nos describiera o que fuera importante para sí mismos. Recuerdo que quedé en blanco, sentí que no tenía nada, nada inspirador. Desde ese momento, para aliviar el vacío, escribí varias en un papel, sin ni siquiera convencerme, hasta que surgió una que deseo plasmar en este trabajo, frase que me motivó en las diferentes fases de la maestría y en este proyecto en particular;

"Hagamos de la vida una buena historia que contar"

Agradecimientos

A mi madre, a mi padre, a mis hermanos y a mis sobrinos, que son el motor de mi vida. A mi novia, quien me acompañó en esta aventura de inicio a fin, a mis amigos, que siempre han estado ahí, en esta y en otros retos que me ha puesto la vida. A la Universidad Nacional de Colombia, que ha sido mi casa y mi mayor orgullo, de quien seré su estudiante por siempre. A la profesora Olga Rodríguez, un pilar dentro de mi formación personal, académica, y profesional a quien siempre le estaré agradecido.

Resumen

Declive en el rendimiento en comprensión de lectura en cuarto grado escolar: un fenómeno presente

Diferentes estudios han puesto en evidencia cómo los resultados en el área de lenguaje y en la comprensión de lectura muestran una tendencia a la baja en el cuarto grado de básica primaria, un fenómeno conocido como declive de cuarto grado, etapa de tránsito entre aprender a leer y leer para aprender. Esta investigación tiene por interés obtener información sobre la presencia de este fenómeno en estudiantes de Bogotá. Se analizaron los reportes de notas de los estudiantes de las asignaturas asociadas a Lengua Castellana, entre 1° y 9° grado escolar para los años 2016 a 2021 de 271 establecimientos educativos, distribuidos en 508 sedes y conformadas por 654,642 estudiantes con base en lo reportado en el Sistema de Apoyo Escolar de la Secretaría de Educación de Bogotá, herramienta que reúne los informes académicos de los estudiantes de algunos colegios de Bogotá. Los resultados indican que hay diferencias estadísticamente significativas entre los grados y que hay un declive en el rendimiento desde los primeros grados hasta el cuarto grado, manteniéndose estable para los siguientes grados, pero con resultados tendientes al promedio-bajo. Se concluye sobre como los resultados apoyan la literatura en el área y cómo resulta importante que los entes institucionales, los docentes, las familias y los estudiantes reconozcan el fenómeno como primer paso para establecer planes de mejora tanto para su reconocimiento como para su intervención.

Palabras clave: Declive de cuarto grado, educación básica primaria en Colombia, comprensión lectora, rendimiento académico, Sistema de Apoyo Escolar SAE, evaluación masiva.

Abstract

Fourth grade reading slump: a present phenomenon

Different studies have shown how the results in the area of language and reading comprehension expose a downward trend in the fourth grade of elementary school, a phenomenon known as *fourth grade slump*, a transition stage between learning to read and reading to learn. This research is interested in obtaining information on the presence of this phenomenon in students in Bogotá. The grade reports of the students of the subjects associated with the Spanish Language, between 1st and 9th grade for the years 2016 to 2021 of 271 educational establishments, distributed in 508 locations and made up of 654,642 students based on what was reported, were analyzed. In the School Support System of the Secretary of Education of Bogotá, a tool that gathers the academic reports of the students of some schools in Bogotá. The results indicate that there are statistically significant differences between the grades and that there is a decline in performance from the first grades to the fourth grade, remaining stable for the following grades, but with results tending to average-low. It is concluded on how the results support the literature in the area and how it is important that institutional entities, teachers, families and students recognize the phenomenon as a first step to establish improvement plans both for its recognition and for its intervention.

Keywords: Fourth grade slump, basic primary education in Colombia, reading comprehension, academic performance, SAE School Support System, Mass evaluation.

Tabla de contenido

Introducción 11

Planteamiento del problema 16

Marco teórico 19

Informe Coleman: Estructura social como elemento transversal al rendimiento escolar 20

Causas de declive de cuarto Grado 27

El rol de la evaluación masiva en el análisis de declive de cuarto grado 38

Metodología 41

Objetivo general 41

Objetivos específicos 41

Procesamiento de la información 41

Población 51

Diseño de investigación 55

Análisis de datos 57

Supuestos estadísticos 59

Resultados 68

Discusión y conclusiones 79

Lista de figuras

Figura 1. Asignaturas relacionadas con el área de lenguaje 42

Figura 2. Distribución de outliers con DE entre 30 y 35. 47

Figura 3. Distribución de outliers con DE entre 35 y 40. 48

Figura 4. Distribución de la variable evaluación escalada 50

Figura 5. Comparación distribución original (Izquierda) frente a la escalada (Derecha) 51

Figura 6. Distribución de la base del Sistema de Apoyo Escolar por año. 52

Figura 7. Distribución SAE por grado escolar. 52

Figura 8. Niveles de desempeño. 53

Figura 9. Distribución del índice de pobreza multidimensional para las sedes educativas del SAE. 54

Figura 10. Distribución de los datos SAE en formato wide. 61

Figura 11. El poder de las pruebas Anova y Krustal Wallis en función de diferencias entre las medias para varios tamaños de muestra en el caso de una distribución normal. 63

Figura 12. El poder de las pruebas Anova y Krustal Wallis en función de diferencias entre las medias para varios tamaños de muestra en el caso de una distribución no normal. 64

Figura 13. Distribución de los grados en la variable escalada evaluación en Lenguaje. 68

Figura 14. Distribución de la agrupación del grupo 1 para los grados 3, 4 y 5 70

Figura 15. Distribución de la agrupación del grupo 2 para los grados 2, 3 y 4. 71

Figura 16. Distribución de la agrupación del grupo 3 para los grados 1, 2, 3 y 4. 72

Figura 17. Distribución de la agrupación del grupo 4 para los grados 2, 3, 4 y 5. 73

Figura 18. Distribución de la agrupación del grupo 5 para los grados 3, 4, 5 y 6. 74

Figura 19. Distribución de la agrupación del grupo 6 para los grados 1, 2, 3, 4 y 5....75

Figura 20. Distribución de la agrupación del grupo 7 para los grados 2, 3, 4, 5 y 6....77

Lista de tablas

Tabla 1. Escalas de calificación de los datos ingresados en SAE.....44

Tabla 2. Niveles de desempeño MEN y mínimo y máximo de los puntajes asignados a estas categorías.....44

Tabla 3. Descripción de los cuartiles para el proceso de estandarización.....46

Tabla 4. Criterios de interpretación del Eta parcial para la estimación del tamaño del efecto en Anova de Medidas Repetidas.....67

Tabla 5. Grupos con mayor cantidad de datos para la realización de Anova.....69

Introducción

El informe de resultados de PIRLS 2011 realizado por el ICFES (2012) reconoce dos elementos importantes al abordar la problemática de lectura en Colombia; el primero, que la lectura es la destreza más importante que un niño puede desarrollar, y por otro lado, que el cuarto grado es un punto de transición determinante en el desarrollo de los niños como lectores, porque en esta etapa la mayoría *ha aprendido a leer y ahora leen para aprender*; los estudiantes encuentran, por medio de la lectura, nuevas palabras, frases e idiomas que mejoran sus capacidades de lenguaje y su vocabulario influyendo sobre el aprendizaje de modelos y conexiones, lo que, a su vez, favorece el desarrollo de habilidades de pensamiento y creatividad. (ICFES, 2012)

Así, el Declive de Cuarto Grado (DCG) o (*Fourth-Grade Slump*) es un fenómeno descrito como aquella etapa de tránsito entre aprender a leer para luego utilizar la lectura como medio para acceder al conocimiento, que fue identificado entre los años 50´s y 60´s por una investigación dirigida por Walter Loban publicada en 1976. El investigador realizó un estudio longitudinal recolectando las notas de los estudiantes por 12 años desde primer grado hasta el grado 12. Encontró que el rendimiento en lectura mostraba unas leves caídas entre el primer y tercer grado, sin embargo, el cuarto grado escolar hacía las veces de embudo, haciendo que la brecha en habilidades de lectura fuese más notoria a partir de este grado y resultaran ampliándose a lo largo del curso escolar (Delsing, 2010). Su investigación mostró como durante los tres primeros años de escuela hubo pequeñas diferencias entre el estudiante con buenos resultados académicos y el estudiante con dificultades. Luego, en cuarto grado, la brecha de habilidades aumentó notoriamente de modo que la diferencia entre estos continuó ampliándose.

En esta línea, la literatura ha reconocido un declive en el rendimiento académico en comprensión de lectura en los últimos grados de primaria especialmente en el cuarto grado y con mayor incidencia en los estudiantes de menores ingresos socioeconómicos. Particularmente, los estudiantes muestran un declive en su rendimiento escolar, en particular en las materias relacionadas con el área de lenguaje debido a que se encuentra en un periodo de tránsito entre aprender a leer, hasta tercer grado escolar, y leer para aprender, pues la lectura se convierte en una herramienta para acceder al conocimiento. Es en ese periodo de tránsito donde se evidencia este fenómeno.

El estudio pionero y con mayor impacto académico sobre el declive de cuarto grado fue desarrollado por los investigadores Chall, Jacobs y Baldwin de la universidad de Harvard en 1990. En su estudio demostraron que los niños tienen un declive en el rendimiento lector alrededor del cuarto grado escolar, y que el fenómeno es más preponderante en estudiantes de bajos recursos. A partir de estos resultados nace una preocupación y una ola de investigaciones, vigente dentro del marco de la política educativa de este país y del mundo, pues los resultados internacionales de alfabetización sugieren que el DCG no es únicamente un fenómeno de EE. UU. (Mullis, Martin, González y Kennedy, 2003; OCDE, 2001).

El estudio mencionado de Chall, Jacops & Baldwin (1990) evaluó inicialmente a 30 estudiantes de los grados segundo, cuarto y sexto, y luego a estos mismos estudiantes un año después cuando estaban en tercero, quinto y séptimo respectivamente. Tomaron aquellos centros educativos que participaban en un programa de alimentación escolar para familias vulnerables de modo que sus participantes se caracterizaban por tener bajos ingresos económicos. En el proceso de recolección de datos usaron pruebas estandarizadas

y las notas de clase de los años en curso. Como resultados encontraron que; aunque los estudiantes de segundo y tercer grado mantenían puntuaciones constantes, los estudiantes de cuarto grado comenzaron a perder terreno rápido estando por debajo de los niveles de lectura esperada. En general, el desarrollo de la lectura en los grados segundo y tercero era fuerte, seguido por la desaceleración a partir de alrededor de cuatro y quinto grado que continuó a través de los grados sexto y séptimo.

Estos estudiantes de bajos ingresos en los grados 2° y 3° lograron tan buenos resultados como los niños de la población normativa en las seis subpruebas del test de evaluaciones diagnósticas de lectura (DAR por sus siglas en ingles). Sin embargo, algunos de los puntajes de los estudiantes de estas escuelas comenzaron a disminuir su rendimiento alrededor del grado cuarto. Desde ese entonces, a esta desaceleración se le conoce como *Fourth Grade Slump* o Declive de Cuarto Grado (DCG). Así, según Chall & Jacops (1983), el cuarto grado escolar se caracteriza porque es en este punto de tránsito donde los niños utilizan la lectura como medio para acceder al conocimiento (Chall & Jacops, 1983). Así, desde el modelo de lectura planteado por estos autores, la lectura es un proceso por etapas que avanza de manera acumulativa, la etapa inicial o prelectura corresponde a los procesos anteriores a la escolarización, las etapas siguientes, uno y dos, se caracterizan por la decodificación de palabras donde van adquiriendo fluidez con el tiempo para luego leer textos sencillos y familiares con un lenguaje cotidiano y propio de su experiencia y capacidad. En su conjunto son conocidas como las etapas de "*aprender a leer"* y tienen su curso entre los grados primero, segundo y tercero.

La tercera etapa está caracterizada por un enfrentamiento a un tipo de textos más complejos en términos lingüísticos y cognitivos ya que el objetivo de la lectura es que

funcione como una herramienta para aprender, es decir, *"leer para aprender".* En este sentido, los textos de cuarto grado contienen nuevas palabras e ideas más allá de la experiencia propia de su desarrollo, se vuelven más variados, complejos y desafiantes, lingüística y cognitivamente, contienen nuevas ideas e implican una exigencia mayor de vocabulario, conocimiento y de pensamiento crítico. Así, el fenómeno de DCG se presenta en los estudiantes que les cuesta hacer la transición de la etapa 2 a la 3, donde la comprensión lectora se ve afectada. (Chall & Jacops, 2003) (Bogaerds-Hazenberg, Evers-Vermeul, & van den Bergh, 2020).

Menciona Donaldson (2010) que desde la última década de los noventa (1990 a 2000) los investigadores tienen suficiente evidencia sobre la presencia del declive en el rendimiento en los últimos grados de la escuela primaria, por tanto, desde entonces, se ha dado paso a nuevas investigaciones sobre las causas, más que a demostrar su prevalencia (Donaldson, 2010).

Sin embargo, las investigaciones en América Latina y particularmente en Colombia resultan escasas frente a la prevalencia de este fenómeno, razón por la que resulta de gran interés investigativo dar cuenta de su prevalencia y de las características particulares del declive, a propósito de las condiciones educativas y sociales del país, en particular las de su capital, pues Bogotá posee características interesantes sobre los demás municipios y ciudades del país tales como: la diversidad en la población en edad escolar y la variabilidad en instituciones educativas, "instituciones que tienen un rendimiento académico excelente, conocidos como escuelas eficaces, y otros con un rendimiento bastante inferior, llamados escuelas ineficaces." (Carreño, 2020, pág. 12), En particular, se cuenta con dos variables en puntos extremos que resultan ser explicativas y de interés al hablar del declive del cuarto

grado; un escenario con grandes desigualdades en cuanto a los niveles de ingreso socioeconómico y una población escolar de bajos resultados en pruebas de rendimiento lector tanto en pruebas nacionales como en pruebas internacionales y en los diferentes niveles de escolaridad.

Planteamiento del problema

Estudiar el declive de cuarto grado (DCG) en Colombia es importante en atención a que es un tema de interés investigativo actual y se ha estudiado en diferentes países y con resultados variados, adicionalmente, en la literatura se evidencia una especial conexión con la política pública, particularmente en Estados Unidos; por ejemplo, Hirsch (2003) menciona que este fenómeno ha sido uno de los mayores fracasos en las escuelas públicas estadounidenses y desde entonces, muchos estudios se han llevado a cabo para investigar las causas del declive y las soluciones para su superación o adaptación. (Delsing, 2020; Hazenberg, Vermeul, & Van den Bergh, H. 2020; Hirsch, 2003).

Por otra parte, es relevante investigar sobre esta temática ya que reconocer qué variables y cómo influyen en el rendimiento en comprensión de lectura permitirá diseñar planes de mejora escolar e influir en términos de diseño y planteamiento de política pública de modo que mejoren los resultados en pruebas nacionales e internacionales.

En efecto, a nivel nacional el informe de resultados agregados de la prueba saber 3° y 5° (ICFES, 2018) muestra que alrededor del 50% de los estudiantes se encuentran en el nivel insuficiente y mínimo en las pruebas saber 3° y 5°, porcentaje que no ha variado desde el 2012 al 2017 y con mayor incidencia según el nivel socioeconómico (NSE). En la misma línea, se encuentran los resultados del Programa para la Evaluación Internacional de Alumnos de la OCDE (PISA), en el que Colombia ha estado por debajo del promedio de los países miembros, incluso en los últimos resultados de 2018 presentó un rendimiento menor al del año 2015 (OCDE, 2019). En los resultados del Estudio Internacional de Progreso en Comprensión Lectora (PIRLS, 2011) que evalúa estudiantes entre 9 y 10 años

de cuarto grado escolar, el 1% de los estudiantes colombianos alcanzaron el nivel avanzado en competencia lectora, y el 28% no alcanzó los niveles mínimos.

Otro elemento a tener en cuenta es que el cuarto grado de primaria en Colombia es uno de los años donde más disminución en matrícula se presenta: "frente al año 2018, los grados que registraron crecimiento fueron primero (1,9%) y segundo (1,4%), con 24.658 más matriculados, los demás grados de básica primaria registraron disminuciones en las matrículas, siendo el de mayor variación el grado cuarto, en -1,0%." (DANE, 2019).

Por otro lado, gran parte de los estudios se han desarrollado en Estados Unidos y Europa, sin embargo, a nivel de América Latina pocos mencionan directamente al declive, particularmente al buscar resultados sobre *fourth grade slump* tanto en Colombia como en la región, las investigaciones son escasas, en parte por tratarse de un término en ingles poco tratado en la literatura en español, pues hablar de declive de cuarto grado es una traducción propuesta por esta investigación buscando conservar el significado de una "*caída en el rendimiento"* o *"un declive en el rendimiento en comprensión de lectura en el cuarto grado escolar",* título de este documento.

No obstante, las investigaciones en la región, que tratan de manera tangencial la temática, hacen referencia al rendimiento académico en el área de lenguaje, continuamente enmarcado en los análisis de evaluaciones internacionales cómo PIRSL, PISA, y algunos abordajes desde la UNESCO que ponen de presente los bajos resultados en la región, donde;

> "(…) la mayoría de los estudiantes realizan una comprensión fragmentaria de los textos que leen, reconocen las palabras de un texto, pero no consiguen determinar por qué se dice lo que se dice o para qué. Es decir, a los niños se les enseña a traducir

las palabras escritas al lenguaje oral, pero sin entender ni interpretar el significado de lo que leen. Se aprende a leer un texto en voz alta o a pronunciarlo, pero no a aprender leyendo" (ICFES, 2001, párr. 3)

Otro elemento de interés es la diversidad en la orientación de las investigaciones en cuanto a las explicaciones del fenómeno y el abordaje del mismo, por ejemplo, algunos estudios como el desarrollado por Etmanskie Partanen & Siegel (2009) con estudiantes canadienses, evidencian que es posible que los porcentajes se sobreestimen y que el declive sea solo eso, un declive. En general, acorde con esta investigación, "los niños parecen recuperarse de su declive cuando se adaptan a las mayores demandas curriculares requeridas desde cuarto grado y más allá". (Etmanskie & Cols 2009, pág. 45).

Sumado a ello, otro estudio con estudiantes turcos muestra que, aun cuando algunos niños presentan inconvenientes en el cuarto grado, esto difícilmente puede atribuirse al estancamiento en su desarrollo del lenguaje o al fenómeno de DCG. Es más probable que las actividades escolares no permitan que estos niños se beneficien de sus recursos lingüísticos completos y que no sea una cuestión individual como lo plantean Chall & Jacobs (1983). (Spindler, Jørgensen & Holmen, 2014)

Acorde con lo presentado, la investigación planteada resulta de interés de cara a las problemáticas en el rendimiento académico en el área de lenguaje en Colombia y la región en razón a que permite aportar en la comprensión del fenómeno incluyendo variables individuales, dando peso al contexto, y teniendo en cuenta las características del sistema educativo, de las familias y de los estudiantes colombianos.

Marco teórico

Las investigaciones en torno al Declive de Cuarto Grado se han centrado en dar cuenta de las *causas* a partir de procesos propios de las variables implicadas en el desarrollo del lenguaje y de los procesos de lectura. Se han definido ampliamente aquellas variables de los estudiantes que están relacionadas con la presencia del declive, elementos que se profundizarán en el apartado *causas del declive de cuarto grado.* No obstante, esta investigación propone una mirada adicional, al reconocer el papel de los centros educativos y de las variables de proceso, individuales, escolares, institucionales, entre otras; modelo que es propio de la línea de investigación en eficacia escolar, siendo una de las áreas de interés conocer las variables que inciden en el rendimiento escolar y por qué ciertos centros tienen mejores resultados que otros, visión que, sumada a la investigación desarrollada alrededor de las causas del declive, permitirá una comprensión global del fenómeno en Colombia.

En efecto, este es el objetivo del movimiento de escuelas eficaces, brindar elementos que resultan importantes a la hora de establecer diferencias en el rendimiento académico, (nivel macro) hasta esas diferencias dadas a nivel micro o de proceso e indagar por cuáles son las variables asociadas al rendimiento. Así, en la investigación en educación sobre el rendimiento escolar ha sido de gran interés preguntarse por aquellas variables que hacen que los centros educativos tengan un mejor desempeño, bajo la lógica que, al conocer las causas, se podrá modificar la realidad y conseguir elevar los niveles de calidad educativa. (Torrecilla, 2003).

Aun cuando esta ha sido una preocupación histórica, es hasta los años sesenta del siglo XX, luego del informe Coleman, cuando se constituye como línea de investigación la

eficacia escolar, que tendrá como objetivo el conocimiento de los factores escolares asociados al logro educativo. Por tanto, este documento describirá inicialmente un apartado sobre el informe Coleman, conectando luego con el movimiento de eficacia escolar, su historia y sus principales aportes de cara al fenómeno en interés. Luego de ello, se abordarán las causas que la literatura ha descrito cómo explicación a este fenómeno para continuar con un apartado que tratará sobre el papel de la evaluación masiva frente al declive, pues parte de la investigación en el área ha tomado evidencia del declive a partir de evaluaciones masivas nacionales e internacionales.

Informe Coleman: Estructura social como elemento transversal al rendimiento escolar

El declive de cuarto grado resulta en una problemática multifacética que retoma varios de los problemas actuales en educación. Por un lado, pone presente como las diferencias en cuanto a los niveles de ingreso socioeconómico se relacionan con el rendimiento académico, tendencia creciente en el grado cuarto de primaria, particularmente en el área de lenguaje y con incidencia en todas la áreas y a lo largo del avance escolar, en el sentido en que la estructura social limita muchos de los procesos implicados en un desarrollo adecuado del proceso lector, pues este es un elemento esencial dentro del funcionamiento del ser humano, ya que la lectura es la forma en cómo se accede al conocimiento del mundo, es un elemento crucial para la interacción social y el aprendizaje. Es por ello por lo que, aprender a leer resulta tan importante como aprender a caminar, hablar, u otra actividad esencial (Dehaene, 2019).

Así, para la Organización para la Cooperación y el Desarrollo Económico (OCDE) las destrezas lectoras son la base del aprendizaje a lo largo de la vida y aumentan las

oportunidades futuras de empleo e ingresos, en consecuencia, los alumnos cuyo rendimiento es más bajo seguramente serán los que tengan menos probabilidad de obtener un empleo prometedor lo cual afecta su movilidad económica. Esto supone una pérdida, no sólo para las personas, sino también para las sociedades dado que su crecimiento económico depende de la contribución de todos sus miembros, (García, 2016; OCDE, 2010).

A nivel global, al analizar los resultados de las pruebas PISA en el área de lenguaje frente al nivel de desigualdad[1] de los países participantes, se encuentra que los países desiguales económicamente tienen resultados por debajo de la media de la OCDE mientras que los países de menor desigualdad obtienen mejores resultados en el área de lenguaje de la Prueba PISA, panorama complejo teniendo en consideración que a nivel global , el índice Gini viene aumentando desde los años 80´s y a la par frente a las desigualdades en el rendimiento escolar (García, 2016)

En efecto, el fenómeno de Declive de Cuarto Grad*o* se describe como un periodo crítico donde los resultados académicos presentan una caída con mayor incidencia en los estudiantes con mayores limitaciones económicas. Así, esta asociación entre rendimiento escolar y nivel socioeconómico tiene un largo desarrollo y constituye en una de las ideas impulsoras y centrales de la investigación en educación escolar, además de constituirse como elemento central del papel de la educación en las sociedades modernas (López, 2018).

Así, para los inicios del siglo XX se potencializa el interés de los países occidentales en la universalización de la educación, conocido como fenómeno de expansión educativa,

[1] Desigualdad descrita en este estudio por el índice Gini: este es un estadístico en economía que indica la desigualdad entre los ingresos de una población.

prolongando los años a los que se tiene acceso, las áreas de conocimiento, buscando un acceso equitativo y universal por parte de la población, comprendiendo la escuela como mecanismo de reproducción social y de generación de oportunidades, sin embargo, la ferviente economía capitalista y globalizada llevaba a grandes desigualdades educativas según los niveles de ingreso de los centros educativos (López, 2018)

En esta línea, para mediados de los 60´s, en Virginia Estados Unidos, se celebra la Conferencia Internacional sobre la Crisis Mundial de la Educación centrando esfuerzos en la necesidad de reformas educativas que permitieran un mayor acceso, principalmente a la secundaria, y una calidad que no estuviese enteramente mediada por el nivel de ingresos de las instituciones, como menciona Marqués, (2016)

> "los diferenciales de acceso a la educación en sus etapas medias y superiores por clase social, (como escuelas independientes para blancos y negros), eran el producto de las condiciones materiales disímiles de los centros en los que se formaban a los alumnos, la propuesta de política pública para el cambio trataba de contrabalancear la báscula de las inversiones educativas, la cuestión radicaba en cómo asegurar la igualdad de oportunidades en una sociedad desigual y la respuesta parecía hallarse en el aumento de la inversión educativa ya que si los alumnos de padres con desventajas sociales disponen de buenas escuelas para llevar a sus hijos, estos tendrían la posibilidad de progresar socialmente." (Marqués, 2016, pág. 111)

Bajo este escenario, en 1964 surge la publicación del informe Coleman *Equalitity of Educational Opportunity Report,* informe histórico, sin precedentes para su época tanto en Estados Unidos como a nivel mundial dada su amplia muestra de 600.000 estudiantes y

cerca de 4.000 instituciones educativas, y sus implicaciones en la investigación educativa posterior a su publicación. (Carabaza, 2016). Este estudio consistió en determinar la naturaleza y extensión de las desigualdades en las escuelas públicas de Estados Unidos, estuvo financiado por el Congreso de los Estados Unidos en el contexto de implantación de la Ley por los Derechos Civiles de 1964 frente a la lucha contra la injusticia racial y la pobreza. Con éste, se intentaba erradicar las injusticias que se producían aún en los Estados Unidos por razón de origen étnico y racial (Marques, 2016). Particularmente se interesó en cuatro cuestiones principales: la segregación racial en las escuelas, las características de las escuelas, el aprendizaje de los alumnos y las relaciones entre las características de las escuelas (Carabaña, 2016).

Pese a esta gran relevancia histórica, sus resultados fueros desalentadores para académicos y responsables de política pública, pues muchas de las variables consideradas teóricamente relevantes para el rendimiento académico no mostraron una mayor influencia. Por un lado, la encuesta encontró mucha segregación racial. Pero, contra lo que se esperaba, no encontró muchas diferencias de recursos materiales y humanos entre las escuelas, por mucho que estuvieran segregadas. Tampoco encontró grandes diferencias de aprendizaje entre las escuelas, y, además, “la mayor parte de las que encontró no se debían a los recursos de los centros, sino a su composición social” (Carabaña, 2016, pág. 10). Los hallazgos más relevantes de este informe, descritos en (Carabaña, 2016), fueron:

La gran importancia del origen familiar para el rendimiento y el hecho de que esta relación no disminuye con los años de escuela. Adicionalmente, la porción relativamente pequeña de varianza entre escuelas no explicada por el origen familiar, indicador del pequeño efecto independiente de las variaciones en las instalaciones escolares, el

currículum en el rendimiento, evidenciando una pequeña porción de varianza en el rendimiento explicada por las variaciones en instalaciones y currículum, y dado el hecho de que no hay factores escolares que den cuenta de mucha variación en el aprendizaje, las características de los profesores explican más que ningún otro. (Coleman; 1966; Carabaña, 2016)

Respecto al componente del lenguaje, elemento de interés de este estudio, indicadores como recursos económicos del hogar, estudios de los padres, interés del alumno en la escuela, explicarían cerca del 30% de la desigualdad en la prueba de aptitud verbal en los cursos 6º, 9º y 12º, mientras que, cinco variables propias de los centros como equipamiento, currículum, calidad y actitudes de los profesores y características del cuerpo estudiantil, explicarían en conjunto menos de un 10% de varianza adicional. En efecto, las medidas usuales de calidad educativa consideradas como recursos para el aprendizaje (gasto por alumno, educación para el profesorado, tamaño de la biblioteca del colegio, entre otros) mostraban poca relación con el logro escolar cuando los estudiantes de bagaje similar eran comparados, mientras que las diferencias entre los distintos bagajes familiares mostraban una relación importante con el rendimiento educativo" (Coleman, 1996, p. 19; Carabaza, 2016 pág. 14).

En este sentido, tras eliminar la influencia del origen étnico y la situación socioeconómica del alumno, las características medidas explicaban apenas un 10% de la varianza de rendimiento de los alumnos, así, este informe concluyó que: "el rendimiento escolar en los Estados Unidos estaba influenciado en gran medida por las características socioeconómicas de los estudiantes, y que las variables asociadas a la institución educativa tenían poco o ningún efecto sobre las diferencias en el desempeño escolar". (Gómez,

Gutiérrez, & Hassan, 2011, pág. 3). Como menciona Marques (2016) "los buenos resultados escolares se alcanzaban cuando los padres de los alumnos eran profesionales y directivos y no cuando eran obreros, independientemente de si sus escuelas estaban más y mejor dotadas, dejando en duda la gran inversión que se estaba promulgando" (Marques, 2016, pág. 113).

Por tanto, considerando el impacto político, económico y social de estos resultados, la posibilidad de transformación y mejora de los centros se vio limitada para este entonces, pues la conclusión general del informe indicaba que los centros importaban poco, ya que la idea de invertir recursos a las escuelas, influyendo sobre variables como estructura, formación de los docentes, mejores pedagogías, entre otras variables propias de las instituciones, no generaría grandes cambios. Como lo menciona el mismo informe;

> "las escuelas son notablemente similares en el modo como se relacionan con el rendimiento de sus alumnos cuando se tiene en cuenta el origen socioeconómico de estos. Es sabido que los factores socioeconómicos guardan una fuerte relación con el rendimiento académico. Cuando estos factores se controlan estadísticamente, resulta que las diferencias entre escuelas dan cuenta de solo una pequeña fracción de las diferencias en el rendimiento de los alumnos" (Coleman, 1966, pág. 22).

En esta línea, invertir en los centros parecía una idea contraria a lo evidenciado por los datos, donde lo que realmente mostraba influencia eran variables que poco se podían modificar, tales como el nivel de ingresos de las familias, nivel de estudio de los padres o constitución de las familias, elementos donde la política pública que se planteaba poco podía hacer, al menos en el corto plazo.

Por tanto, surgía como producto de este informe, una época gris y nublada de la investigación en educación, pues la idea de que el rendimiento académico está altamente influenciado por las condiciones socioeconómicas se convertiría en una regla de oro que apoyarían tanto los que aportaban elementos en línea con este informe como los que buscaban controvertirlo, pues más allá de la discutida influencia de los centros, el peso de las variables propias de las condiciones sociales en el rendimiento escolar se afirmaron mucho más con los datos de este informe, tesis que ya venía cobrando fuerza en la investigación sobre el declive de cuarto grado, donde la influencia de la variable de origen social frente al rendimiento era la principal explicación a la hora de comprender a qué estudiantes perjudicaba con mayor fuerza esta etapa de transición.

Años siguientes, los bajos resultados escolares para finales de siglo XX, evidentes en pruebas estatales en Estados Unidos como la ISAT (The Illinois State Achievement Test) sumado a lo descrito en el Informe Coleman, generaron un gran debate político y académico sobre las causas de bajo rendimiento, discusión que llegó hasta el congreso de los Estados Unidos siendo una gran influencia en la ley *No Child Left Behind Act (NCLB)* (*Que ningún niño se quede atrá*s). Esta ley fue aprobada por el congreso de los Estados Unidos en el gobierno del presidente George W. Bush en el año 2002 y autorizó a varios programas educativos federales para que fueran administrados por los estados, los cuales deberían evaluar a los estudiantes en lectura y matemáticas en los grados 3 a 8 y una vez en la escuela secundaria se esperaría que todos los estudiantes alcancen o excedan los estándares estatales en lectura y matemáticas. El enfoque principal fue cerrar las brechas en los logros de los estudiantes, brechas que resultaron evidentes con la investigación sobre el fenómeno de declive de cuarto grado y los resultados del informe Coleman. Así, esta

normativa buscaba brindarles a todos los niños una oportunidad justa, equitativa y significativa de obtener una educación de alta calidad (Washington Office of Superintendent of Public Intruction, 2022).

Sin embargo, estudios posteriores a su implementación indican que la NCLB impactó aún más en el logro estudiantil, haciendo más evidente el declive de cuarto grado (Aronson, 2004; Delpit, 1996; Landsman, 2004; Delsing, 2010). Así, en el 2005, el 44% de los alumnos de tercer grado de una escuela de Chicago cumplieron o superaron el nivel de grado en lectura, pero en 2006, solo el 37% de los alumnos de cuarto grado alcanzaron el grado nivel o superior, una caída del 7% del rendimiento en lectura. Del mismo modo, una revisión de los resultados de las pruebas de 2006 de los alumnos de tercer grado indicó que el 59 % obtuvo un puntaje a nivel de grado o superior en lectura. Un año más tarde, los puntajes de cuarto grado mostraron que el 41% obtuvo el grado nivel o superior en lectura, una disminución del 18% (Junta de Educación del Estado de Illinois, 2007; Delsing, 2010).

Causas de declive de cuarto Grado

Un primer elemento al hablar de las causas del declive hace referencia a que los estilos de enseñanza de los maestros cambian durante los años de la escuela primaria, resultando en una menor interacción social. En el jardín de infancia, los estudiantes pasan mucho tiempo trabajando en grupos y socializando, y cuando los estudiantes pasan a cuarto grado realizan menos actividades de este tipo. Daniels (2002) considera que la interacción social en entornos grupales puede ser una forma poderosa de aumentar las actitudes de lectura. Debido a que los maestros de grados superiores utilizan instrucción y

tienden a descuidar los aspectos sociales de la lectura, las actitudes hacia la lectura disminuyen.

Otro factor que potencialmente podría causar este declive es la enseñanza orientada para dar cumplimento a unos estándares educativos. Como se mencionó anteriormente, la brecha entre los estudiantes de bajo rendimiento y los estudiantes de alto rendimiento aumentaron desde que se implementó la Ley *No Child Left Behind* en Estados Unidos en el gobierno de George Bush (NCLB). "Normatividad que para algunos profundizó y es una de las causas del declive de cuarto grado en Estados Unidos" (Delsing, 2010, pág. 13)

Frente a ello, Houston (2007) menciona los llamados *"siete pecados capitales del la NCBL"*, el primero fue asumir que las escuelas estaban rotas cuando, de hecho, no lo estaban asumiendo que los resultados de los test estandarizados daban cuenta del todo del funcionamiento del centro, dejando de lado aspectos como la colaboración, el ingenio, la resolución de problemas y otras menos tangibles. Así para el segundo, que menciona que la NCLB confunde *Testing* con *education* y deja por fuera todas las variables de interés que una prueba estandarizada no logra recolectar. El tercero menciona que los programas de fortalecimiento no tienen en cuenta que si a los estudiantes de las escuelas de bajo rendimiento se les instruye continuamente en las habilidades básicas, mientras que a los estudiantes en una estructura de clase media participan en un plan de estudios más rico, esto amplía aún más la brecha de logros entre los estudiantes de nivel de pobreza y de clase media, lo que lleva que a una misma actividad los estudiantes no estén igual de preparados dadas sus diferencias de ingreso y de posibilidades de estimulación mental.

El cuarto pecado está relacionado en que el miedo y la coerción proporciona una motivación negativa y, a su vez, traumatiza a los niños. En la medida en que se dedica más tiempo a aprender cómo aprobar un examen y menos tiempo se gasta en instrucción. Además, una mayor presión para pasar la prueba puede causar mayor ansiedad para los estudiantes al tomar la prueba y esta puede ser una fuente de bajo rendimiento en pruebas estandarizadas,

El quinto menciona que los padres y las comunidades no entienden fácilmente la NCLB, no fueron tenidos en cuenta y lo les resulto claro por ejemplo porque niños que hablan inglés nativo frente a otros tomaban un mismo examen, o por qué también lo presentaban estudiantes con necesidades especiales, entre otras comparativas. De igual modo el sexto aspecto, pues los educadores no fueron parte del proceso de creación del lineamiento, lo que se traduce en desconexión, poca afinidad con el mismo y bajos resultados. Finalmente, el último punto menciona que las pruebas estandarizadas podrían sofocar el pensamiento creativo y bajar el nivel del rendimiento académico, pues deja de lado otras habilidades importantes dentro del desempeño mismo. (Houston, 2007; Delsing, 2010)

En esta línea, Springen (2007) cree que la enseñanza para la prueba da como resultado que los maestros descuiden el uso de múltiples géneros para involucrar a los estudiantes. Además, este estilo de enseñanza puede resultar en que los maestros no permitan elección para los estudiantes, que es un poderoso motivador y mejora las actitudes.

Otro elemento explicativo en la investigación sobre el declive es la capacidad del lector para reconocer e integrar las relaciones significativas entre los elementos de un texto tales como los eventos, los hechos y los escenarios (van den Broek y Kremer, 2000). Los lectores construyen una representación de la base de texto (proposiciones que se derivan directamente del texto a nivel de oración) pero no comprenden cómo se organizan estas proposiciones a nivel global.

Según el modelo construcción-integración (Kintsch, 1988 , 2013 ; van Dijk & Kintsch, 1983), es precisamente esta reorganización de la información proposicional, a nivel global y la sucesiva integración con el conocimiento previo, las que son cruciales para la comprensión del texto. Cuanto mejor se organiza la información en esquemas mentales y se elabora con conocimientos previos relevantes, más coherente es el modelo de situación del texto del lector y, como resultado, su comprensión es mejor (Kintsch, 1988 , 2013 ; Kintsch & van Dijk, 1978; Bogaerds & Cols,2020).

Desde esta perspectiva, el DCG puede ser comprendido como un fenómeno que ocurre cuando los lectores no son sensibles a la organización jerárquica de la información en los textos, lo que dificulta la construcción del modelo de situación apoyado en sus conocimientos previos que sirven como estructura de orden superior que ayuda a predecir y organizar el contenido textual mientras leen (Zwaan, 1994 ; Bogaerds & Cols, 2020) (Kintsch, 2013).

Por otra parte, se ha sugerido que el fenómeno de DCG tiene como superficie la discapacidad de lectura de aparición tardía *(Late-Emerging Reading Disabilities)* (Leach & Scarborough & Rescorla, 2003; Etmanskie, 2009). Este fenómeno describe que las

dificultades de lectura pueden enmascararse en los primeros grados y luego emerger en el proceso de tránsito propio del cuarto grado escolar ya que algunos niños dependen casi por completo de la memorización visual de palabras y, por lo tanto, parecen estar logrando aprender a leer durante los primeros años hasta que esta estrategia se vuelve difícil de manejar. Incluso, "los niños que inicialmente dominan la decodificación con éxito podrían encontrar dificultades cuando las palabras reconocidas y deletreadas se vuelven más complejas fonológica y morfológicamente después de los grados primarios." (Leach & Cols, 2003, pág. 214).

Por tanto, "es en el proceso de decodificación donde los estudiantes de este grado escolar se suelen enfrentar con palabras más complejas y con sílabas que resultan desconocidas, lo que traduce en una debilidad y limitación de la capacidad cognitiva y de memoria para facilitar el reconocimiento de palabras." (Etmanskie, 2009, pág. 4) (Leach & Cols, 2003). En esta línea, "la decodificación de lectura no es suficiente y limita la comprensión de lectura directamente, ya que se ve influida capacidad de aplicar las reglas de correspondencia de letras y sonidos al leer palabras y no palabras". (Perfetti, 1985; Lyon, 2002; Vellutino, 2003; Best, Floyd, McNamara, 2004 pág. 2)-

Estas caídas en el rendimiento lector pueden atribuirse a diferentes variables, una de ellas constituye el tipo de material dispuesto para que los niños aprendan a leer, caracterizados por un contenido limitado, básico, débil tanto en la construcción del vocabulario como en el conocimiento de dominio necesario para avanzar a libros de texto

de nivel superior[2]. Por ejemplo, "un libro de ciencias es diferente en formato y contenido a un texto de matemáticas y para conseguir el logro académico los estudiantes deben adaptarse a textos académicos variados, facilitando la amplitud de sus habilidades de alfabetización." (Etmanskie, 2009, pág. 3)

Según López (2015) los libros de texto utilizados en la parte inicial de la primaria son predominantemente narrativos y poco expositivos, estos últimos caracterizados por su dificultad, ya que no siempre tienen la misma estructura, presentan gran cantidad de información nueva e incluso suelen incluir algún tecnicismo. Todo ello va a dificultar la comprensión, evidenciando mayores problemas para realizar las inferencias que para responder a preguntas literales, por lo tanto, parece que la capacidad para deducir una información que no aparece en el texto es a veces limitada, ya que los libros de textos presentan la mayoría de las veces preguntas literales por lo que los estudiantes desarrollan unas estrategias de búsqueda rápida de la información sin necesitar una comprensión más general, por lo tanto, este tipo de estrategia no funciona para responder a preguntas inferenciales que requieren una comprensión mucho más profunda.

En este orden, esta orientación hacia la literalidad y la exclusión de cuestiones valorativas de los libros de texto no promueve un pensamiento crítico ni procesos de metacomprensión claves en el desarrollo de una buena compresión lectora en razón de que las actividades incluidas en los libros son demasiado obvias y no dan lugar al pensamiento crítico porque la mayor parte se pueden responder sin apenas haber leído el texto o a partir

[2] Los *"basal readers"* a los que se hace referencia corresponden a los libros de texto que son utilizados para enseñar la lectura en el contexto escolar.

de ideas literales (Chrobak, 2000; Puente 1994: López, 2017). Así, de acuerdo con esta perspectiva, “este material fomenta que los alumnos no sean capaces de desarrollar capacidades como la deducción o cómo exponer la discrepancia de ideas o de puntos de vista”. (López, 2015, pág. 448-449).

Relacionado con lo anterior, el género del texto es otra variable relacionada con el declive en el rendimiento lector dado que estos grados escolares enfrentan desafíos de comprensión crítica hacia los textos expositivos, particularmente aquellos que cubren material científico. Los textos expositivos, contrario a los narrativos, están caracterizados por ser poco familiares para los estudiantes de primaria, mientras que los textos narrativos tienen como meta comprender y relacionarse con conjuntos de eventos, personajes y escenarios conocidos cercanos a su ambiente y su contexto. Así, la diferencia principal entre los textos narrativos y expositivos se refiere a sus demandas de conocimiento. De hecho, el punto clave en el rendimiento lector en el cuarto grado es que los lectores que tienen un mayor conocimiento sobre el tema del texto tienen mejor memoria para el texto y desarrollan una comprensión más profunda de la información transmitida (Best, Floy & Mcnamara, 2004).

En efecto, para la comprensión del texto narrativo, propio de etapas iniciales como se mencionaba anteriormente, la capacidad de decodificación es la fuerza motriz que implica un proceso de decodificación de palabras eficiente y automático que puede liberar espacio en la memoria de trabajo, centrando está en la construcción de un modelo de situación a partir de su conocimiento de la vida y sus experiencias educativas. Sin embargo, la construcción de este modelo de situación, al enfrentarse a un texto de género expositivo o científico, no basta con el conocimiento de la vida y las experiencias, se hace necesario un

conocimiento previo sobre el tema del cual se está leyendo (Best, Floy & Mcnamara, 2004). Esto implica que los niños entre los grados tercero y cuarto de primaria con menos conocimientos previos tienen dificultades para formar un modelo de situación coherente al leer un texto expositivo porque no son capaces de generar las inferencias necesarias, al no poseer los conocimientos previos específicos del tema que leen. (Best, Floy & Mcnamara, 2004). (Mcnamara, Ozuru & Floyd, 2011).

Del mismo modo, la cohesión de los textos es otro elemento importante dentro del declive de cuarto grado. Este efecto del conocimiento previo sobre la comprensión lectora está regulado por la forma en que se escriben los materiales de lectura, en especial en las pistas y señales que ayudan a los lectores a relacionar la información distribuida en diferentes partes del texto garantizando su cohesión.[3] "Se considera que los textos tienen poca cohesión cuando la construcción de una representación coherente a partir del texto requiere muchas inferencias basadas en el conocimiento del lector. Los textos se consideran de alta cohesión cuando los elementos dentro del texto proporcionan pistas más explícitas sobre las relaciones dentro y a través de oraciones" (McNamara, Louwerse, McCarthy y Graesser, 2010: McNamara, Ozuru & Floyd, 2011 pág. 232).

Como tal, los textos utilizados en el cuarto grado escolar, caracterizados por su baja cohesión, imponen mayores demandas de procesamiento al lector, en especial para los lectores que no cuentan con un conocimiento previo del tema del cual se van enfrentar, pues este conocimiento faltante le dificultará llenar esos vacíos productos de la baja

[3] Un ejemplo de las señales a las que se hace referencia son los conectores lógicos que explícitamente muestran la relación entre dos oraciones; por ejemplo, por tanto, sin embargo, pero, aunque, entre otras.

cohesión, limitando la integración de significados de oraciones individuales en una representación coherente de situaciones o eventos representados por el texto. En otras palabras, “la baja cohesión en lectores con bajo conocimiento dificulta la construcción de un modelo de situación" (Kintsch, 1988, 1998: McNamara, Ozuru & Floyd, 2011 pág 232).

Relacionado con lo anterior, otra variable importante al hablar del declive de cuarto grado es el papel del vocabulario o el conocimiento de palabras, elemento clave en el desarrollo de una buena comprensión lectora. (Hirsch, 2007). Así, los niños que carecen de conocimiento de vocabulario probablemente tendrán dificultades para comprender textos como los expositivos debido a que no tienen una experiencia sustancial con recursos informativos y vocabulario relacionado, por tanto, es menos probable que hayan adquirido las habilidades necesarias para comprensión de este género de texto (Gregg y Sekeres 2006: Sanacore & Palumbo, 2008)

Esta adquisición temprana de vocabulario está mediada por algunas características de los estudiantes, entre ellas el nivel socioeconómico. El estudio clásico en el área realizado por Hart y Risley (1995) reporta que los cuidadores con un Nivel Socio Económico (NSE) más alto tienden a hablar más con sus hijos, en el transcurso de una semana los niños con NSE más alto escucharon casi cuatro veces más palabras que los niños con NSE más bajo, igualmente los hogares de escasos recursos exponen a los niños a estructuras de oración mucho más simples, una brecha que se mantiene constante durante los primeros tres años. Por tanto, “un alumno de primer grado con alto rendimiento conoce cerca del doble de palabras que un alumno de bajo rendimiento y, a medida que estos estudiantes avanzan de curso, el diferencial aumenta”. (Hirsch, 2007, pág, 239).

Según Nagy y Anderson (1984); Etmanskie (2009), los lectores con dificultades en los grados intermedios de primaria leen 100.000 palabras por año en comparación con 10.000.000 palabras por año para los lectores típicos, generando una brecha que crece de manera exponencial con el pasar del tiempo, y generando lo que Stanovich (1986) identificó como el "*efecto Matthew",* donde los estudiantes que son buenos lectores leerán mejor, tendrán más vocabulario, más conocimientos y mejores habilidades lingüísticas mientras que los lectores deficientes no suelen interesarse en la lectura y esa baja motivación se traduce en menor interacción con los textos, menor vocabulario, menor conocimiento y pocas habilidades de lectura.

Por tanto, una variable transversal es el papel de la motivación dentro del declive de cuarto grado donde la motivación hacia la lectura suele ser baja trayendo como consecuencia un menor acercamiento y un crecimiento exponencial de las limitaciones en la comprensión lectora (Muñoz, Ferreira, Sánchez, Santander, Pérez & Valenzuela, 2012).

En particular, un aspecto clave entre los buenos lectores y los que presentan dificultades se refiere a las actitudes hacia la lectura. Frente a ello, la investigación clásica de McKenna, Kear y Ellsworth (1995), ha estudiado esa variable en relación con su influencia en el declive. Mencionan que la relación más fuerte está entre las actitudes de cuarto grado hacia la lectura recreativa y los puntajes de rendimiento posteriores, pues los puntajes más bajos de rendimiento en lectura se correlacionan con actitudes de lectura más bajas. (Donaldson, 2010) (McKenna, Kear y Ellsworth,1995),

Este estudio empleó la Encuesta de Actitud de Lectura Elemental (ERAS) para determinar una relación entre actitudes lectoras y otros componentes de la lectura. La

investigación encontró una correlación positiva entre las actitudes de lectura y el rendimiento en lectura a propósito del declive encontrando que se ve afectado por la motivación, la actitud y la auto percepción del logro. (McKenna, Kear y Ellsworth, 1995).

Un elemento que retoman estos autores frente al rendimiento lector como elemento diferenciador son las actitudes hacia la lectura académica y la lectura recreativa, pues la lectura recreativa, especialmente la lectura de ficción resulta en un indicador del rendimiento en la competencia lectora. Lógicamente, los estudiantes que leen fuera de la escuela aplicarán las habilidades y estrategias de lectura con más frecuencia que los niños que no leen fuera de la escuela, aumentando su capacidad lectora. En este sentido. los estudiantes que tienen más probabilidades de leer fuera de la escuela son estudiantes con una actitud positiva hacia la lectura recreativa. (Ogle & Cols, 2003; Donaldson, 2010)

Así, los estudiantes deben tener una actitud lectora positiva para lograr altos niveles de lectura al enfrentar el declive, pues en este caso las actitudes lectoras influyen en las autopercepciones del estudiante sobre sus propias habilidades lectoras. Según Gerber, Ginsberg & Reiff (1992); Donalson (2010), cuando un estudiante toma una la postura personal de que él o ella va a tratar de obtener un mejor resultado, lo más probable es que obtenga resultados ajustados a ello, por tanto, antes de intentar convertirse en un lector exitoso, el estudiante debe creer que puede tener éxito y este camino se puede orientar con las lecturas recreativas como anclaje, en razón de que una actitud negativa hacia la lectura resulta en un grave obstáculo para este proceso. (Gerber, Ginsberg & Reiff, 1992; Donaldson, 2010),

Finalmente, otras variables relacionadas con la comprensión de lectura en el cuarto grado escolar son la fluidez y la memoria de trabajo. Así, un lector experto ha automatizado muchos de los procesos relacionados con la lectura, libera memoria de trabajo y la ocupa en el proceso interpretativo. A diferencia de un novato, cuando comprende un texto integra los conocimientos y el vocabulario previo con el que cuenta junto con el material nuevo que se le presenta y construye un modelo coherente del significado del texto, rápidamente hace conexiones agregando múltiples componentes en un solo *"chunk"*[4], que a su vez se aloja en un campo independiente de la memoria de trabajo disponiendo de una red de información latente que facilita el proceso inferencial del material nuevo. Mientras que el lector novato, con menor conocimiento previo y menor cantidad de palabras en su repertorio, le tomará más tiempo ya que no podrá hacer conexiones con los contenidos nuevos y su memoria de trabajo se verá ocupada en procesos subyacentes como la decodificación. (Hirsch, 2007)

El rol de la evaluación masiva en el análisis de declive de cuarto grado

Otro tema que atañe al fenómeno de declive de cuarto grado es el papel de la evaluación en la educación, en particular la evaluación masiva nacional e internacional y su vínculo con la toma de decisiones y la generación de política pública entorno a los sistemas educativos. Gran parte de los estudios sobre esta temática parten de un proceso de recolección de información a partir de pruebas estandarizadas tanto locales como

[4] Chunking es definido por George A. Miller como la manera en que las personas reúnen múltiples componentes en un solo ítem, el que a su vez se aloja en un surco independiente de la memoria inmediata. "El gato está sobre el felpudo" es una oración que se puede recordar fácilmente, y los lectores expertos pueden reproducir fácilmente las 24 letras, no porque las letras se recuerden individualmente, sino porque la oración se recuerda como un trozo (chunk) cuyos sub-elementos pueden reconstruirse a partir de un conocimiento previo del idioma escrito. (Hirsch, 2007)

internacionales, pues son un elemento que por un lado brindan datos sobre la calidad educativa, posibles explicaciones de los resultados y propuestas para mejorar el rendimiento (Hambleton, 2002: Odendahl, 2011)

Una de las pruebas más relevantes en el ámbito de evaluación educativa y del área de lenguaje y directamente relacionada con el declive de cuarto grado es PIRLS (Progress in International Reading Literacy Study). La Asociación Internacional para la Evaluación del Logro Educativo (IEA) lo define de la siguiente forma:

> "Esta prueba se administra en una etapa de transición clave en el desarrollo de la lectura de los niños: el cambio de *aprender a leer a leer para aprender* ya que evaluar el rendimiento en lectura en esta etapa crucial proporciona a los educadores y a los responsables de la formulación de políticas información clave sobre la eficacia de su sistema educativo y ayuda a identificar áreas de mejora. *Es reconocido como el estándar mundial para evaluar las tendencias en el rendimiento en lectura en cuarto grado*. PIRLS proporciona datos comparativos internacionales sobre qué tan bien leen los niños y ofrece información relevante para las políticas para mejorar el aprendizaje y la enseñanza." (Asociación Internacional para la Evaluación del Logro-Educativo (IEA), 2022, par, 7)

El estudio se desarrolla cada 5 años desde el año 2001, y Colombia ha participado en tres oportunidades, 2001, 2006 y 2011. Para esta última participación Colombia obtuvo 448 puntos, resultado por debajo de la media de 500 puntos. Solamente el 1% de los estudiantes colombianos que participaron en PIRLS 2011 alcanzaron el nivel avanzado en competencia lectora; el 9% obtuvo un puntaje clasificado en nivel alto, el 28% en nivel medio, y el 34%

se ubicó en el nivel bajo. El 28% restante no alcanzó los niveles mínimos de competencia lectora (obtuvieron resultados por debajo de 400 puntos), lo cual significa que no han tenido un aprendizaje eficaz el proceso de lectura (ICFES, 2021).

Estos resultados a la luz de esta investigación pueden ser interpretadas como una dificultad propia de lo que implica el declive de cuarto grado, donde las variadas limitaciones de los estudiantes colombianos llevan a que su rendimiento académico en la prueba sea bajo, pues esta etapa de transición del *aprender a leer al leer para aprender* requiere una serie de elementos para desarrollarse de manera acorde, que en el contexto colombiano se traducen en barreras de oportunidad.

Metodología

Objetivo general

Brindar información sobre la presencia del fenómeno de declive en el rendimiento en comprensión de lectura en el cuarto grado (DCG) en los colegios públicos de Bogotá a partir del análisis de la información del Sistema de Apoyo Escolar de la Secretaria de Educación entre los años 2016 y 2021 de los grados primero a sexto.

Objetivos específicos:

- Analizar los puntajes de los informes en el área de lenguaje de los estudiantes de primero, segundo, tercero, cuarto, quinto y sexto con el fin de comprobar el cambio de tendencia en los resultados en el cuarto grado de primaria.
- Describir las tendencias en el rendimiento escolar a través de los datos de SAE de los diferentes grados escolares.
- Abordar cómo se ha definido el declive de cuarto grado, las variables asociadas y su relevancia frente al desempeño lector en Colombia.

Procesamiento de la información

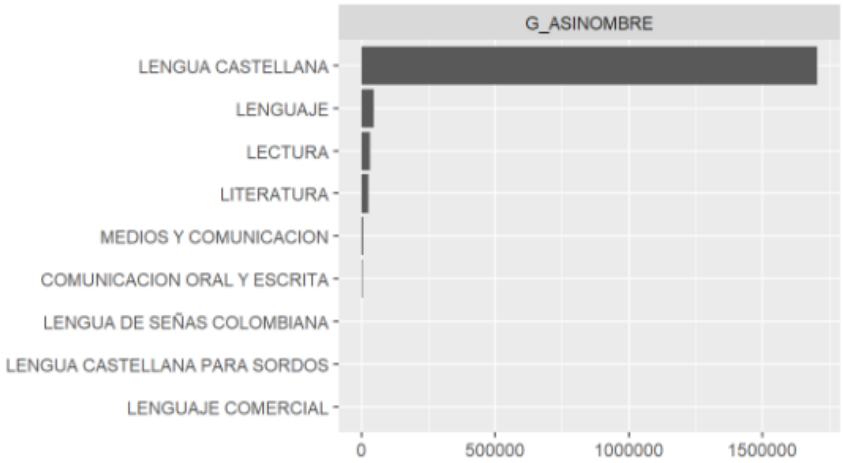
Esta investigación corresponde a un análisis de la información a partir de los datos del Sistema de Apoyo Escolar (SAE), el cual tiene una matriz de datos con registros correspondientes a todas las materias registradas en varios centros educativos de la ciudad de Bogotá.

El SAE es una herramienta creada para brindar apoyo a los colegios en varios procesos institucionales: generación del carnet estudiantil, captura de información de la

evaluación de estudiantes y generación de boletines, entre otros, donde cada institución toma la decisión de adoptar o no el uso de esta herramienta. En el 2016, el sistema contaba con la participación de más de 200 colegios a nivel Bogotá (Secretaría de Educación de Bogotá, 2022). En el SAE cada profesor de cada asignatura ingresa el resultado de la evaluación del estudiante al sistema y cada colegio establece criterios de evaluación en relación con los niveles de desempeño del MEN: Bajo, Básico, Alto, Superior, por tanto, cada colegio establece, de forma autónoma, equivalencias entre los resultados de la evaluación de sus estudiantes y los niveles de desempeño del MEN.

Para la investigación, el primer paso fue realizar la solicitud de la base de datos SAE. La Dirección de Evaluación de la Secretaría de Educación de Bogotá alistó la información acorde con los requerimientos del investigador anonimizando nombres de los estudiantes y de los centros por tratarse de información sensible y de información de menores edad. También implicó seleccionar las materias que guardaban relación con el área de lenguaje, cómo se presenta en la Figura 1.

Figura 1. Asignaturas relacionadas con el área de lenguaje



Fuente: Elaboración propia a partir de los datos de SAE.

Como criterio se seleccionó una nota por año, aun cuando el SAE cuanta con notas por periodo para cada año, debido a que la información entre estudiantes y entre sedes era variable, encontrando inconsistencias en el número de periodos y dificultades técnicas para generar la matriz con esa información. Por otro lado, aun cuando el SAE es una estrategia de la Secretaría de Educación de Bogotá que se viene manejando desde el año 2009, los datos de años anteriores al 2016 no eran comparables por cuestiones propias del sistema,

Luego de conocer las características de la información entregada por la Secretaría de Educación de Bogotá, frente al procesamiento, se realizó la depuración de la información y la limpieza de la base de datos con el paquete R lo que implicó ajustar temas formales como reemplazar puntos por comas, valores en letras por números, categorías mal escritas, entre otras[5], entendiendo que la base entregada consistió en una consulta directa del sistema y que requería este proceso de depuración.

Posteriormente, se eligió la asignatura tener en cuenta para los análisis, como lo muestra la Figura 1, la matriz de datos entregada muestra un total de 9 asignaturas y para algunos estudiantes contaba con registros de varias de ellas al año mientras que para otros estudiantes no, por esta razón, se toma como unidad de análisis los registros con la asignatura “Lengua Castellana”, que representa más del 95% de los datos. Luego de ello, considerando que la matriz entregada describe un registro por año, se depuraron aquellos registros que no cumplían con este criterio, igualmente para los estudiantes que repitieron el mismo grado en diferentes años, dejando únicamente el último registro.

[5] Esta información se puede encontrar en detalle en el script de programación de R.

El último paso en la depuración, y el más sensible de todo el proceso, consistió en un proceso de estandarización de la variable "Evaluación" debido a que la escala de calificación resultaba distinta incluso en el mismo grado, el mismo año, la misma sede y jornada con 6 escalas de calificación distintas como lo muestra la Tabla 1, ello producto de la autonomía de cada docente al escoger la escala y la asignación a las categorías de desempeño MEN.

Tabla 1. Escalas de calificación de los datos ingresados en SAE

Min	**Max**
1	3
0	5
0	10
0	25
0	50
0	100

Fuente: Elaboración propia a partir de los datos de SAE.

Adicionalmente, los criterios de pertenecer a un nivel de desempeño MEN están superpuestos como lo muestra la Tabla 2.

Tabla 2. Niveles de desempeño MEN y mínimo y máximo de los puntajes asignados a estas categorías

Mínimo – máximo	**Nivel de desempeño MEN**
70 a 100	SUPERIOR
70 a 100	ALTO
0 a 90	BASICO
0 a 73	BAJO

Fuente: Elaboración propia a partir de los datos de SAE.

En efecto, una característica del SAE es que cada colegio establece criterios de evaluación en relación con los niveles de desempeño del MEN: Bajo, Básico, Alto, Superior, igualmente cada colegio establece, de forma autónoma, equivalencias entre los resultados de la evaluación de sus estudiantes y los niveles de desempeño del MEN, razón que explica esta superposición entre la asignación categorial no excluyente de los valores numéricos a cada nivel de desempeño MEN.

Estas características de la información de la matriz del SAE, lleva a que sea necesario un proceso de estandarización que permita la comparación de los puntajes dentro de una misma escala, pues las comparaciones entre puntajes sin que resulten dentro de una misma escala de calificación llevan a interpretaciones erróneas y desajustadas. Por tanto, el proceso de estandarización consistió en determinar los puntajes que fueran atípicos para luego realizar un proceso de estandarización para todos los registros de modo que quedaran en una misma escala.

El punto de partida es que una nota registrada ha de tener la misma escala de calificación si pertenece a una misma jordana, a una misma sede, a un mismo año y a un mismo curso, sin embargo, la matriz de datos mostraba casos extremos que, aun bajo esta condición, tenían datos atípicos, por ejemplo, notas de 4, 4,2, 4,7, 3 en una misma jornada-curso- año-sede con un dato atípico de una nota o registro de 78.

Por ello, un primer paso consistió en estimar estos datos atípicos *"outliers"*, considerando el primer cuartil, la mediana, tercer cuartil, el rango intercuartílico y la

desviación estándar de cada registro agrupado una misma jornada-curso- año-sede, como lo ejemplifica la Tabla 3.

Tabla 3. Descripción de los cuartiles para el proceso de estandarización.

AÑO	SEDE	JORNADA	GRADO	MIN	Q1	MEDIANA	Q3	MAX	MAX_RIC	MIN_RIC	DE
2020	S0203	TARDE	5	4	4.000	4.00	4.000	4	4.000	4.000	NA
2020	S0343	MAÑANA	1	2	2.750	4.00	5.000	5	8.375	-0.625	1.2817
2021	S0148	TARDE	3	1	2.000	2.00	4.000	4	7.000	-1.000	1.0267
2021	S0263	TARDE	3	1	1.000	1.00	1.000	1	1.000	1.000	0.0000
2018	S0048	TARDE	4	3	6.095	7.03	7.665	95	10.020	3.740	14.0371

Fuente: Elaboración propia a partir de los datos de SAE.

Con esta información, un primer criterio de depuración fue crear un rango intercuartílico amplio que permita la conservación de los datos atípicos reales, para luego poder estimar, a partir de la desviación estándar, aquellos datos que se alejan de la distribución y que serán señalados como *outliers* y depurados de la matriz a analizar.

Para definir hasta qué punto la Desviación Estándar (DE) resulta un buen estimador de los datos *outliers* se generaron cinco muestras al azar para ver cómo se distribuían los datos conforme a diversos rangos la DE, encontrando un mayor ajuste cuando DE es menor a 35 como se puede evidenciar la Figura 2 y 3.

Figura 2. Distribución de outliers con DE entre 30 y 35.

```
## La llave es  2021 S0008 TARDE 6

##  La DE es  30.42

##  Y los valores son  10 20 10 10 20 10 70 90 60 60 30 20 80 70 83 70 73 60
20 91 10 10 10 10 63 10 100 40 70 40 87 20 60 94 60 70 10 30 10 60 75 67[1] "
"

## La llave es  2021 S0038 MAÑANA 1

##  La DE es  30.019

##  Y los valores son  85 60 10 75 90 74 90 35 30 10 80 70 10 45 84 25 85 60
10 80 80 80 15 10 90 80 80 35 35 95 50 95 10 84 85 90 80 90 80 70 30[1] "
"

## La llave es  2021 S0056 TARDE 6

##  La DE es  31.647

##  Y los valores son  45 27 18 81 18 30 78 42 72 90 78 96 68 18 0 54 73 44 92
70 66 66 0 66 88 78 80 78 0 81 68 0 85 81 30 67 0 18 0 70 18 85[1] "
"

## La llave es  2021 S0008 TARDE 6

##  La DE es  30.42

##  Y los valores son  10 20 10 10 20 10 70 90 60 60 30 20 80 70 83 70 73 60
20 91 10 10 10 10 63 10 100 40 70 40 87 20 60 94 60 70 10 30 10 60 75 67[1] "
"

## La llave es  2021 S0038 MAÑANA 1

##  La DE es  30.019

##  Y los valores son  85 60 10 75 90 74 90 35 30 10 80 70 10 45 84 25 85 60
10 80 80 80 15 10 90 80 80 35 35 95 50 95 10 84 85 90 80 90 80 70 30[1] "
```

Fuente: Elaboración propia a partir de los datos de SAE.

Sin embargo, cuando la DE es mayor a 35, se comienzan a presentar *outliers* (resaltados con negrita) como se puede evidenciar en la Figura 3.

Figura 3. Distribución de outliers con DE entre 35 y 40.

```
## La llave es  2018 S0383 TARDE 4

##  La DE es  39.662

##  Y los valores son  3.7 3.6 3.9 3.9 3.7 90 90 3.6 4 4 3.6 4.4 4.2 3.6 92
3.5 3.7 3.9 96 92 3.5 3.6 3.5 4 4 4 90 96 3.5 3.8 3.5 4.2 4.2 4 4.3 92 3.5 3.5
90 96 90[1]

## La llave es  2016 S0201 TARDE 2

##  La DE es  35.586

##  Y los valores son  3.43 3.23 3.1 86 3.98 3.03 3.88 3.83 3.73 3.6 85 3.65
3.88 3.35 3.95 3.18 3.5 3.33 3.75 82 3.95 85 82 3.55 2.8 84 3.83 82 85 3.73
3.63 3.45 3.65 83 3.28[1] "                                                 "

## La llave es  2018 S0201 TARDE 2

##  La DE es  37.71

##  Y los valores son  3.9 3.07 3.07 3.07 3.13 3.73 3.87 86 84 3.7 3.07 3.37
3.47 3.8 86 3.8 80 96 88 82 3.77 3.1 3.77 3.13 3.93 3.87 3.13 80[1] "
"

## La llave es  2018 S0383 MAÑANA 4

##  La DE es  39.566

##  Y los valores son  3.5 3.7 3.8 3.5 90 3.5 3.6 4 3.8 3.6 3.5 3.7 3.9 100
3.5 100 4 3.9 90 3.7 4 3.8 3.8 3.8 90 3.5 3.8 4 90 100 90 3.8 3.6 90 90 3.5
3.7 4 90 3.7 3.7 100 3.5 3.8 4 3.8 3.6[1] "
"

## La llave es  2018 S0201 TARDE 2

##  La DE es  37.71

##  Y los valores son  3.9 3.07 3.07 3.07 3.13 3.73 3.87 86 84 3.7 3.07 3.37
3.47 3.8 86 3.8 80 96 88 82 3.77 3.1 3.77 3.13 3.93 3.87 3.13 80[1] "
```

Fuente: Elaboración propia a partir de los datos de SAE.

Como se puede observar, cuando las DE son mayores a 35 la probabilidad de que al interior de esos datos haya inconsistencias, es alta. Por tanto, se determinó como criterio de exclusión DE por encima de 35.

Luego de depurar los valores atípicos, el paso siguiente fue determinar una misma escala para los datos, un proceso de estandarización. Frente a ello, se decidió optar por la estandarización MIN MAX, ya que resulta ser conservadora frente a la distribución original de los datos, comparada por ejemplo con una estandarización Z que afecta en mayor medida la distribución de los datos.

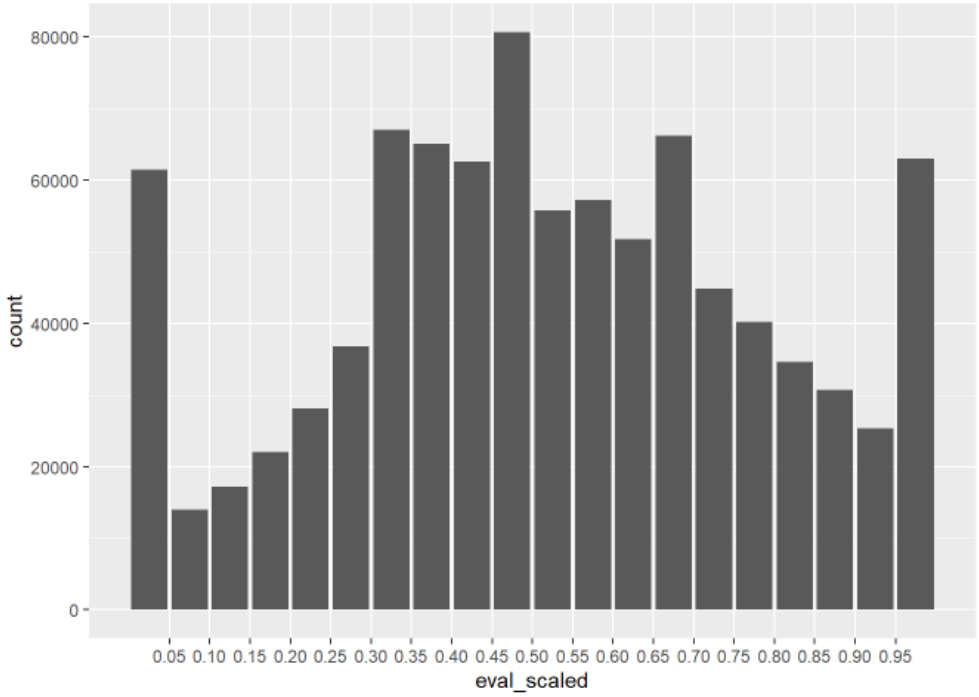
Este método realiza una transformación lineal de los datos originales V en el intervalo especificado [*newmin, newmax*] y esta definió como:

V' = (V - min) * (newmax - newmin) / (max - min) + newmin

La ventaja de este método es que preserva exactamente todas las relaciones entre los datos. No introduce ningún potencial sesgo en los datos. La desventaja es que se encontrará un error “fuera del límite” ("out of bounds “) si un futuro ingreso de datos cae fuera del rango original, sin embargo, como el rango de los datos es fijo, de 0 a 100, resulta en un estadístico robusto y apropiado para las particularidades del estudio. (Acuna, SF,)

Un punto importante que mencionar es que la distribución, tanto de los datos originales como los estandarizados, tiene valores de alta frecuencia en los extremos de la distribución, como lo muestra la Figura 4. Se puede notar que el cero (0) y el uno (1) son valores extremos que se encuentran entre los datos con mayor frecuencia, esto debido posiblemente a las características de asignación de notas de los docentes, que suelen usar estos valores con bastante regularidad. Ello implica dos cosas, por un lado, que es una razón importante para que el estadístico elegido sea la trasformación lineal *minmax,* pues estos valores extremos cero (0) y uno (1) no se ven afectados y no tienden a la normalización. Por otro lado, esta particularidad afecta los criterios de normalidad bastante comunes en el análisis estadístico, pues aun cuando la teoría central del límite define que cuando crece el número de sumandos, se asegura su convergencia hacia una distribución normal, para este caso los valores extremos ya nombrados no lo permiten. Por tanto, será un elemento de análisis retomado con profundidad en el apartado correspondiente de este documento.

Figura 4. Distribución de la variable evaluación escalada



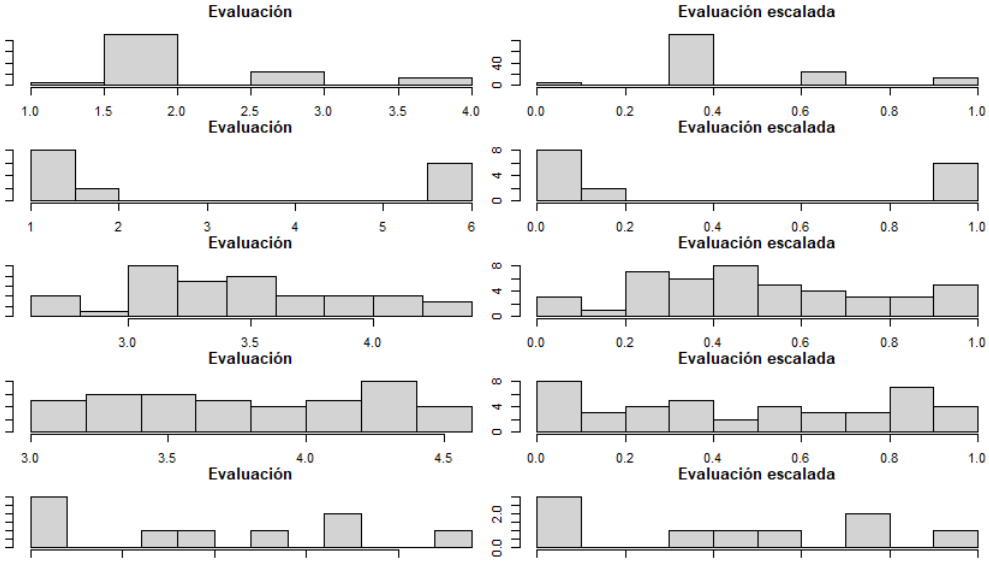
Fuente: Elaboración propia a partir de los datos de SAE.

De igual modo, una consecuencia de la estandarización es que cuando en un mismo grupo todos obtienen la misma nota, la estandarización de ese grupo da como resultado un error. Estos datos, los que no tienen varianza al interior de cada grupo, se han depurado, así como estudiantes que sólo tienen una calificación porque no permiten hacer comparaciones.

Finalmente, como evidencia de la equivalencia entre la puntuación original versus la trasformada, la Figura 5 muestra la distribución de los datos en las dos condiciones, y se puede notar que la distribución es similar en las dos condiciones, adicional a ello, una segunda validación fue establecida por el coeficiente de correlación de Pearson, que resulta

una medida independiente de las escalas de las variables, que indica la relación lineal entre dos variables aleatorias, indicado una alta correlación entre las dos estimaciones.

Figura 5. Comparación distribución original (Izquierda) frente a la escalada (Derecha)

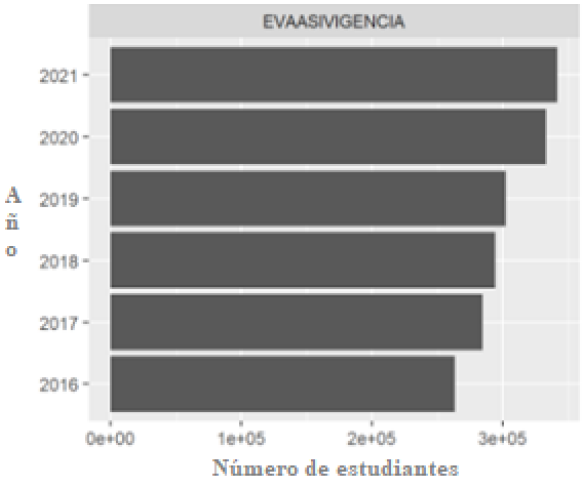


Fuente: Elaboración propia a partir de los datos de SAE.

Población

La matriz de datos contó con 1,819,416 registros con un total 271 establecimientos educativos, distribuidas en 508 sedes y conformadas por 654,642 estudiantes desde el año 2016 hasta el año 2021. La distribución de registros por año se presenta en la Figura 6, se evidencia que a medida que aumentan los años también aumenta el número de registros de la base.

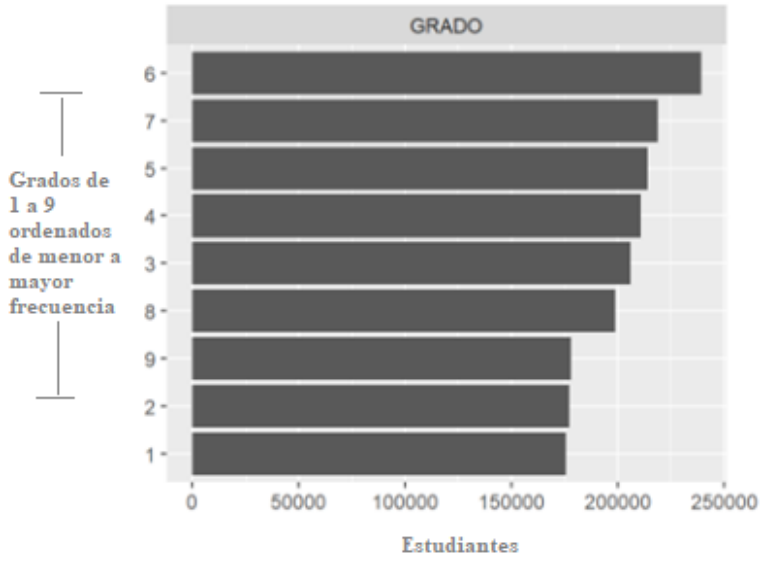
Figura 6. Distribución de la base del Sistema de Apoyo Escolar por año.



Fuente: Elaboración propia a partir de los datos del SAE.

En cuanto a la distribución por grado, descrita en la Figura 7, cada grado tiene datos representativos de al menos 175.000 datos para primero, segundo y noveno y con un valor de mayor de 250.000 datos para el grado sexto.

Figura 7. Distribución SAE por grado escolar.

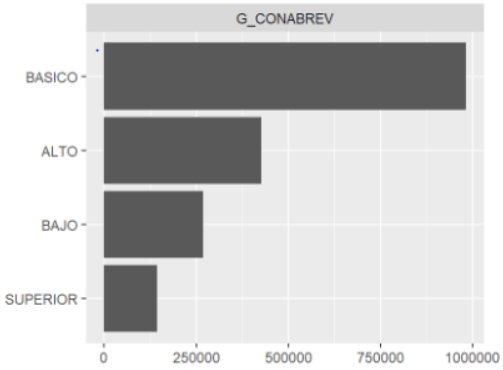


Fuente: Elaboración propia a partir de los datos del SAE.

En cuanto a la variable género, la muestra está distribuida en 52% hombres y 48% mujeres. Respecto a la jornada, los datos presentan mayor presencia de estudiantes de jornada mañana con un 48%, un 42% para la jornada tarde y un menor número de datos para la jornada única con 7% y completa un 2%. Frente a la ubicación de los centros educativos, estos son particularmente urbanos, característica importante de los colegios en Bogotá con un 95%, con una mínima presencia de colegios rurales 3% y de expansión[6] con un 2%.

Por otra parte, la base dispuesta describe una variable que define el nivel de desempeño de la evaluación siguiendo las categorías del nivel de desempeño del Ministerio de Educación Nacional (MEN), en 4 niveles crecientes, *Bajo, Básico, Alto y Superior* descritos en la Figura 8, donde se puede apreciar que el nivel de mayor frecuencia es el Básico, seguido por el Alto, Bajo y luego el Superior respectivamente.

Figura 8. Niveles de desempeño.

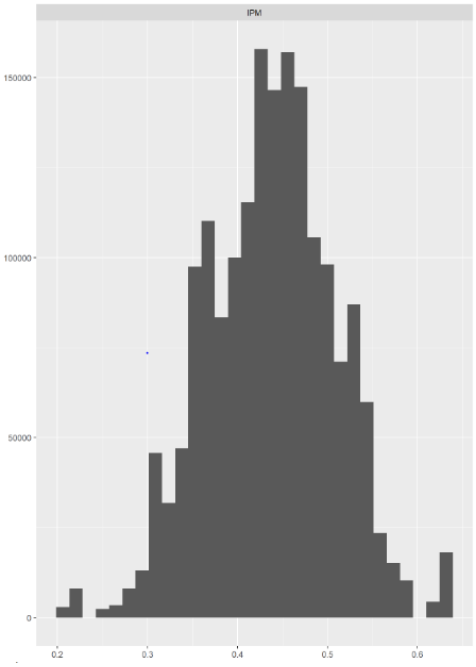


Fuente: Elaboración propia a partir de los datos del SAE.

[6] Los colegios de expansión son aquellos que pertenecen a zonas alejadas de la urbanidad, pero no pertenecen aun a la ruralidad.

Finalmente, la base entrega información sobre el Índice de Pobreza Multidimensional de cada sede educativa, este es un indicador que refleja el grado de privación de las personas en un conjunto de dimensiones. La medida permite determinar la naturaleza de la privación (de acuerdo con las dimensiones seleccionadas) y la intensidad de esta. su distribución se describe en la Figura 9 y se puede notar que tiende a la izquierda la distribución de los datos, consistente con el registro de este índice para el año 2021, que fue del 18,1 % (Red de Pobreza Multidimensional, 2022).

Figura 9. Distribución del índice de pobreza multidimensional para las sedes educativas del SAE.



Fuente: Elaboración propia a partir de los datos del SAE.

Esta información del IPM en la matriz SAE tiene como punto de partida el trabajo conjunto entre la Secretaría de Educación del Distrito (SED) y el Instituto de Estudios Urbanos (IEU) de la Universidad Nacional de Colombia, dentro de la línea “Caracterización socioeconómica de la matrícula oficial de Bogotá D.C.” Así, el índice calculado para Bogotá

toma como referencia el IPM oficial para Colombia (CONPES, 2012; Angulo, Diaz y Pardo, 2011), y se ajusta a la información a las fuentes: Sisbén y EMB 2014. (Secretaría de Educación de Bogotá, 2022)

Diseño de investigación

Esta investigación puede ser comprendida dentro de las características de un diseño longitudinal de medida repetidas, debido a que se cuenta con al menos dos registros correspondientes a dos cursos de un mismo estudiante, con un máximo de cinco. Así, como señala Balluerka & Vergara (2002) "los diseños de medidas repetidas están caracterizados por el registro de diversas medidas de la variable dependiente en un mismo grupo de sujetos y radica en la posibilidad que éste brinda de registrar un conjunto de puntuaciones o medidas de una variable, en dos o más puntos en el tiempo" (Balluerka & Vergara, 2002, pág, 241). En efecto, en este tipo de diseños el tiempo resulta en una variable de interés y cada registro se convierte en el punto a comparar, y en la medida en que el registro corresponde a un mismo sujeto, la varianza por error es menor si se compara con otro tipo de diseños. "De esta manera, el diseño intrasujeto consigue mayor precisión que cualquier otro tipo de diseño en la estimación de los efectos experimentales. De acuerdo con esta cualidad, diversos autores afirman que los diseños de medidas repetidas poseen mayor potencia estadística que los diseños completamente aleatorios (Arnau & Bono, 2008). Por tanto, desde una mirada estadística Llorca & Delgado (2004) describen que "un estudio longitudinal implica por lo menos más de dos mediciones a lo largo de un seguimiento; y que implica más de dos, ya que todos los estudios de cohortes tienen como mínimo dos medidas, una inicial y la de cierre del seguimiento" (Llorca & Delgado, 2004, pág.145).

Una ventaja importante de este tipo de estudios es que como cada sujeto se considera como su propio control ello aumenta la potencia estadística en razón de que la variabilidad intra-sujeto es menor que la variabilidad entre-sujetos, además de permitir diferenciar los cambios por el efecto de la edad (tiempo) frente al cambio de las cohortes, lo que facilita la comprensión de la variable de interés de un mismo sujeto en varios registros y varios registros en diferentes cohortes o grados, lo que reduce la varianza de error y minimiza la varianza sistemática secundaria (Balluerka & Vergara, 2002).

Otra característica para tener en consideración en este tipo de diseños es que resulta esperable que todo el grupo de sujetos participantes no tengan registros en todas las temporalidades establecidas, para el caso de esta investigación, la movilidad entre colegios, la repitencia, la deserción, el acceso al aplicativo por parte de la institución, entre otros factores, hacen que la base no cuente con todos los registros desde el grado primero hasta el grado noveno

También, un elemento a considerar dentro del diseño experimental de medidas repetidas cuando existe una variable independiente aplicada en las diferentes medidas son los efectos de período y los efectos residuales. El primero hace referencia al sesgo derivado del orden en el que se les administran los tratamientos a los sujetos, ya que puede representar mayor variabilidad si se tiene en cuenta el aprendizaje, la familiaridad, la frustración o la fatiga entre las fases iniciales y finales. Por otra parte, los efectos residuales implican que el efecto de la variable independiente o de tratamiento no ha desaparecido en el periodo de la siguiente medición, lo que conlleva a considerar otra varianza en la interpretación del cambio.

Sin embargo, la presente investigación tiene elementos de un diseño longitudinal de medidas repetidas, pero no hace una manipulación directa de las variables independientes que resultan de interés, razón por la que no hay control del efecto del periodo ni del efecto residual. En este sentido, esta investigación se plantea desde un tipo de diseño no experimental, pues no puede generar grupos de manera discrecional y hacer algún tipo de manipulación de las variables independiente de interés, además la recolección de la información es después del hecho, lo que lo caracteriza dentro del tipo de diseños ex-postfacto, haciendo alusión a que; "primero se produce el hecho y después se analizan las posibles causas y consecuencias, por lo que se trata de un tipo de investigación en donde no se modifica el fenómeno o situación objeto de análisis" (Bernardo, J. y Caldero, J.F., 2000; Gordillo, Mayo, Lara & Gigante, 2010, pág. 3).

Análisis de datos

En lo concerniente al análisis de datos para este tipo de diseños, pueden seguirse diversos procedimientos. Así, "cuando la variable de respuesta se distribuye normalmente, es posible aplicar las técnicas de análisis multivariante, análisis de la variancia de medidas repetidas, análisis de curvas de crecimiento, modelos de efectos mixtos y los modelos de ecuaciones de estimación generalizada." (Liang y Zeger, 1986; Bono & Arnau, 2008, pág., 34)

En términos estadísticos, en un diseño de medidas repetidas se evalúa el efecto de un factor sobre una variable independiente medida en dos o más ocasiones para un mismo individuo o para varios individuos que pertenecen a un mismo grupo (Grados escolares). Como los múltiples valores en la variable dependiente están correlacionados entre ellos ya

no es posible usar ANOVA en su forma tradicional, donde los errores de estimación son independientes entre los niveles del factor. Por tanto, una variante del Anova es el Anova de medidas repetidas, donde cada sujeto se considera un factor. El punto radica en que el factor "sujeto" se considera un factor aleatorio en vez de un factor fijo, ya que un factor fijo es replicable ya que un investigador puede rehacer el estudio usando los mismos niveles del factor, pero en este caso, los sujetos no son replicables, en el sentido que son una muestra de los niveles que un factor puede asumir, como lo es cuando un estudiante X hace grado primero en el año y, el investigador no puede replicar tal cual esa información. Por tanto, "la hipótesis en un ANOVA de medidas es si existen diferencias entre las medias del factor A (i.e., meses) controlando por las diferencias asociadas a cada caso." (Muñoz, 2019. párr. 5)

En este sentido, siguiendo a Llorca & Delgado (2004), el análisis longitudinal se puede utilizar cuando existen mediciones del efecto y/o de la exposición en diferentes momentos del tiempo. Así, se plantea la relación entre una variable dependiente "Y", (en este caso el rendimiento escolar en lenguaje), que es función de una variable "X" que cambia en el tiempo (El declive de cuarto grado) y otra Z que es estable en el tiempo (el mismo estudiante avanzando por los grados escolares), variables que se estudian en N sujetos en K momentos del tiempo, (un número determinado de estudiantes a lo largo del ciclo escolar), lo cual se expresa según la ecuación siguiente:

$$Yit = \beta xit + zi\ \alpha + \varepsilon it$$

Donde; i hace referencia al individuo en un tiempo(t) determinado, mientras Z corresponde a un valor estable, e al error estimado para el individuo(i) y el tiempo(t), el coeficiente β o intercepto es una constante que se denomina punto de cruce, y es el valor

promedio de “y” cuando X es cero (0), Es decir β0 es el rendimiento promedio de los estudiantes cuándo las variables predictoras, en este caso el declive de cuarto grado en función del tiempo toma el valor de cero.

Supuestos estadísticos

El Anova de medidas repetidas proporciona un estadístico F que compara si la varianza explicada es significativamente mayor que la varianza no explicada y resulta útil para dar cuenta de las diferencias entre medias de dos o más medidas. Para efectuar este análisis, se ha de cumplir con una serie de supuesto: que las respuestas entre los distintos sujetos sean independientes; que la distribución de las variables de interés siga una distribución normal y que el conjunto de datos no tenga datos perdidos. (Bono & Arnau, 2008), otros supuestos importantes es que; “la variable independiente debe ser categórica y la variable dependiente debe ser continua; que los grupos deben ser independientes entre sí; que no debería haber valores atípicos significativos; que debería haber homogeneidad de varianza entre los grupos; de otro modo, el valor de p para el estadístico F podría no ser fiable.” (Goss & Meneses, 2018, pág. 74)

Frente a ello hay varios aspectos que han de ser considerados para la interpretación de los resultados de esta investigación, por un lado, la independencia entre las respuestas es un supuesto difícil de cumplir en un estudio longitudinal y mucho más cuando también se trata de un diseño exposfacto. Los estudios de tipo longitudinal de medidas repetidas en investigaciones educativas traen consigo una varianza importante ya que la variable de interés no es del todo libre de influencia entre una medición y otra, pues hay variables como las características de los sujetos, de los centros educativos y variables de otro tipo que

influyen en que la media entre un tiempo y otro puedan variar y no se deban a un único factor. Adicionalmente, los estudios exposfacto no permiten proyectar un diseño desde cero que permita un mayor control de las variables extrañas o intrusas, en estos se parte de una información ya recolectada, en este caso representada por los datos del SAE de la Secretaría de Educación.

Por otra parte, este estudio, al partir de una base institucional, tiene un gran alcance, representatividad y un número alto de registros. Sin embargo, una característica de la distribución de las notas asignadas, descritas en la matriz bajo la variable "Evaluación", es que hay una alta frecuencia de dos valores extremos, los ceros y los unos (ya en la variable escalada), pues resulta frecuente que el docente asigne una nota mínima y una nota máxima y ello repercute directamente en el cumplimiento del supuesto de normalidad, necesario tanto para este tipo de análisis, como para gran parte de los análisis paramétricos. Pese a esta característica, esta investigación asume la normalidad de las distribuciones considerando los descrito en la Teoría Central de Limite. En línea con lo planteado por (Bologna, 2013) que menciona que este supuesto puede reemplazarse por tamaños de muestra suficientemente grandes que permitan la aproximación normal usando el Teorema Central del Límite (Bologna, 2013)

> "Este teorema nos dice que si una muestra es lo bastante grande (generalmente cuando el tamaño muestral (n) supera los 30), sea cual sea la distribución de la media muestral, seguirá aproximadamente una distribución normal. Es decir, dada cualquier variable aleatoria, si extraemos muestras de tamaño n (n>30) y calculamos los promedios muestrales, dichos promedios seguirán una distribución normal. Además, la media será la misma que la de la variable de interés, y la desviación

estándar de la media muestral será aproximadamente el error estándar" (Canal, SF pág. 119)

Otro elemento importante referente a los supuestos es la presencia de datos perdidos y datos faltantes dentro de la cadena de registros. Así, como se mencionó con anterioridad, el SAE brinda información desde el grado primero hasta el grado noveno, sin embargo, cuando se analiza la matriz de datos desde el formato *wide*, (Ver Figura 10) se puede notar que hay muchos valores de los que no se tiene conocimiento, encontrando que, del universo de registros de la matriz, pocos estudiantes tienen una cadena de registros completa, ello implica que la base en este formato tiene un número importante de datos perdidos.

Figura 10. Distribución de los datos SAE en formato wide.

COD_ESTUD	8	9	6	7	5	4	2	1	3	NumeroEvaluaciones
E0521440	0.66667	0.66667	NA	NA	NA	NA	NA	NA	NA	2
E0126237	NA	NA	0.62145	0.62143	0.61579	NA	NA	NA	NA	3
E0270650	NA	NA	NA	NA	0.00000	NA	NA	NA	0.00000	2
E0267775	NA	NA	NA	0.00000	0.60976	0.31788	NA	NA	0.57432	4
E0037950	0.60123	NA	NA	NA	NA	NA	NA	NA	NA	1

Fuente: Elaboración propia a partir de los datos de SAE.

Finalmente, un último elemento importante dentro de los supuestos para el análisis de varianza de medidas repetidas es la presencia extrema de datos atípicos, aspecto relacionado con la distribución de los datos y la característica de asignación de notas ya mencionada. Una posibilidad para ajustar la distribución de los datos a la normalidad es la estandarización Z o Z normalizada, ello haría que la distribución de datos "aplanara" los extremos y se ajustaran los datos conforme a los supuestos requeridos de normalidad y

presencia de datos atípicos. Sin embargo, ello implicaría que la trasformación realizada no fuese representativa de la distribución de las características originales de los datos, por ello, la estandarización min-max empleada es garantía de validez de la información pese a resultar una limitante frente al cumplimiento de supuestos estadísticos para efectuar los análisis correspondientes.

Ahora bien, una decisión considerada es el uso de estrategias no paramétricas, en este caso Kruskal-Wallis, recomendada cuando no se logran los supuestos estadísticos requeridos (Goss, Meneses, 2018) y el criterio está dado por la prueba que represente mayor potencia estadística según las particularidades del estudio.

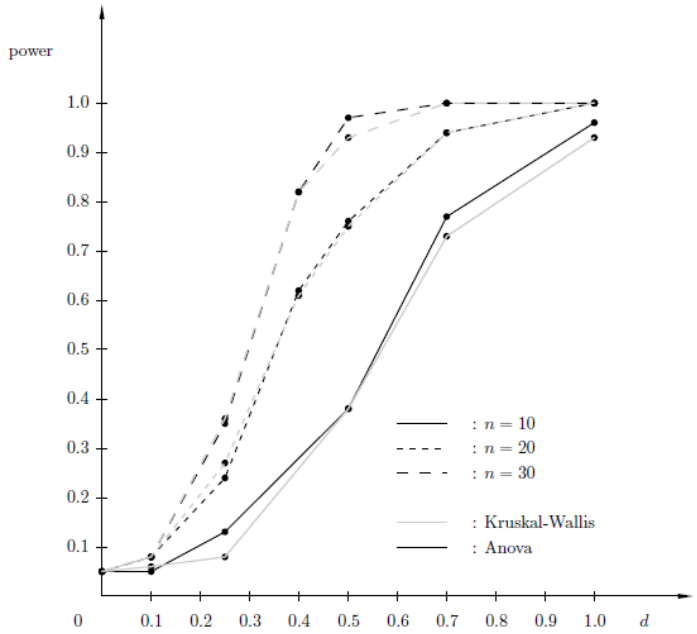
Bajo esta lógica, Hecke (2010), hace una comparación de estas dos pruebas considerando el no cumplimiento del supuesto de normalidad. Por un lado, menciona que la no normalidad tiene algún efecto sobre el error Tipo I, pero el efecto es mínimo cuando las varianzas son de los grupos son iguales, mientras que las probabilidades de cometer error tipo II aumentan en mayor medida, sin embargo, “cuando se cumplen los supuestos de normalidad y de homogeneidad de varianza, las pruebas paramétricas resultan mucho más potentes” (Wilcox, 1995; Hecke, 2010, pág. 1)

No obstante, dado que la prueba Kruskal-Wallis trabaja sobre los rangos de las observaciones y no sobre los datos en sí, ello implica que asume que las observaciones de cada grupo provienen de poblaciones con una misma distribución, y en efecto la potencia se ve reducida si un grupo, (por ejemplo cuarto de primaria) esta sesgado a la izquierda mientras que los primeros grados escolares podría estimarse estarían sesgados a la

derecha, característica que haría que la prueba de Kruskal Wallis pueda dar resultados inexactos en línea con los resultados de Fagerland & Sandvik (2009) ; Hecke (2010).

De igual modo, Hecke (2010) también menciona que cuando la distribución es simétrica el poder de Kruskal Wallis es comparable con su equivalente paramétrico, como se puede evidenciar en la Figura 11.

Figura 11. El poder de las pruebas Anova y Krustal Wallis en función de diferencias entre las medias para varios tamaños de muestra en el caso de una distribución normal.

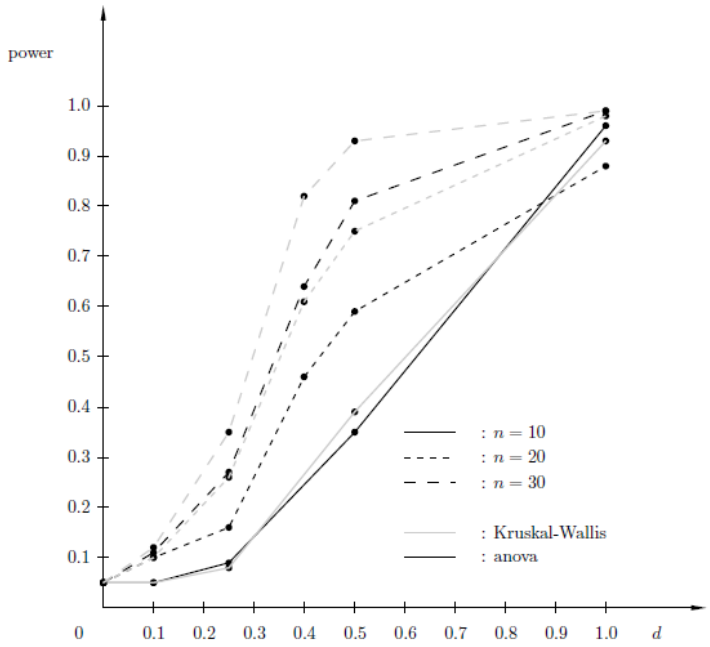


Fuente: Tomado y adaptado de Hecke (2010).

Nota: Prueba de permutación (2500 permutaciones simulación Monte Carlo) para α = 0,05 en el caso de 2500 muestras de las distribuciones elegidas para el Anova y la prueba de Kruskal-Wallis. (Hecke, 2010, pág. 5)

Mientras que "cuando la distribución es asimétrica presenta una pendiente menos pronunciada para las dos pruebas y el rendimiento de la prueba no paramétrica de Kruskal-Wallis da como resultado una mayor potencia a medida que n aumenta". (Hecke, 2010, pág. 6)

Figura 12. El poder de las pruebas Anova y Krustal Wallis en función de diferencias entre las medias para varios tamaños de muestra en el caso de una distribución no normal.



Fuente: Tomado y adaptado de Hecke (2010).

En conclusión, siguiendo Hecke, (2010) aun cuando la prueba no paramétrica puede resultar robusta cuando no se cumplen el supuesto de normalidad y también potente cuando se cumple, hay que considerar que la distribución de los datos resulta más restrictiva en la prueba Krustal Wallis frente al Anova, en la medida en que la primera trabaja en una trasformación de los datos a rangos, suponiendo una misma distribución para que resulten equivalentes, mientras el Anova es menos restrictivo en este aspecto y más fiel a la representatividad de la información para el análisis de los datos del SAE, razón por la que ha elegido la prueba paramétrica por encima de la no paramétrica.

Frente a este aspecto, como menciona Bono & Arnau (2008) “el Anova soporta su validez en que las correlaciones entre los distintos pares de medidas han de ser constantes

y ello lleva a que su uso en diseños longitudinales sea un gran problema dadas las correlaciones entre las observaciones para un mismo sujeto en distintos momentos". (Bono & Arnau, 2008, pág. 36)

Ahora bien, como ya se ha mencionado, este tipo de estudios supone que las varianzas entre cada los niveles de la variable independiente son iguales, en este caso los grados escolares, primero vs segundo, primero vs tercero, primero vs cuarto y así sucesivamente, es decir, si un estudio tiene 3 niveles (A, B y C), la esfericidad asume que las varianzas son iguales (Goss & Meneses, 2018): Por tanto, para su estimación se hace uso del Criterio de esfericidad de Mauchley. (Marín, 2019)

$$\text{Varianza (A-B)} \approx \text{Varianza (A-C)} \approx \text{Varianza (B-C)}$$

Esta pone a prueba la hipótesis nula de que las varianzas de las diferencias son iguales. "En muchos casos, las medidas repetidas violan el supuesto de esfericidad, lo que puede conducir a un error de Tipo I. Cuando este sea el caso, se pueden aplicar correcciones al estadístico F como las correcciones épsilon (ε) de Greenhouse-Geisser y de Huynh-Feldt"(Goss & Meneses, 2018)

Ahora bien, lo anterior respecto a los supuestos, no obstante, luego de ejecutar los Anova se indicará si se presentan o no diferencias significativas entre los grupos que se comparan, sin embargo, es necesario explorar en que pares de agrupaciones se encuentran estas diferencias, este paso es lo que se conoce como pruebas *post hoc,* Goss & Meneses, (2018) mencionan dos alternativas;

a) Bonferroni: que puede ser muy conservador, pero garantiza el control del error de Tipo I a riesgo de reducir la potencia estadística.

b) Holm-Bonferroni: que es un método secuencial Bonferroni menos conservador que el test Bonferroni original, razón por la que será el estimador para esta investigación. (Goss & Meneses, 2018)

Por último, adicional a la significación estadística del Anova y de las pruebas posthoc realizadas, un análisis posterior es la estimación del tamaño del efecto, que nos indica, adicional a la diferencia estadística encontrada, que tan grande resulta esa diferencia. Según Ledesma, Macbeth & Kohan (2008); Cohen (1992) el tamaño del efecto (TE) se define como "el grado de generalidad que posee esa superioridad de A sobre B en la población de la que se obtuvo la muestra estudiada. De esta manera, el TE se refiere a la magnitud de un efecto que es, en este ejemplo, la diferencia entre un tratamiento nuevo A y otro clásico B. Si A es realmente mejor que B, interesa saber en qué medida se espera este fenómeno en la población" (Ledesma, Macbeth & Kohan; Cohen, 1992b, pág. 426)

Menciona Castillo & Bravo (2015), que "la presentación e interpretación del tamaño del efecto como una medida de la significancia práctica de los resultados no es solo un aspecto de sentido común en la investigación con metodología cuantitativa en psicología, sino también un objeto de exigencia de importantes instituciones académicas internacionales, como la American Psychological Association (APA) y la American Educational and Research Association (AERA)". Castillo & Bravo, 2015, pág. 138)

Para su estimación se hará uso del estadístico Eta cuadrado parcial (ηp^2), pues este resuelve el problema de la sobreestimación de la varianza en la población permitiendo la comparación del efecto de la misma variable en distintos estudios, adicional, esta es la forma más habitual de reportar el tamaño del efecto en el ANOVA de medidas repetidas.

Para su interpretación, según Goss & Meneses (2018), se define:

Tabla 4. Criterios de interpretación del Eta parcial para la estimación del tamaño del efecto en Anova de Medidas Repetidas.

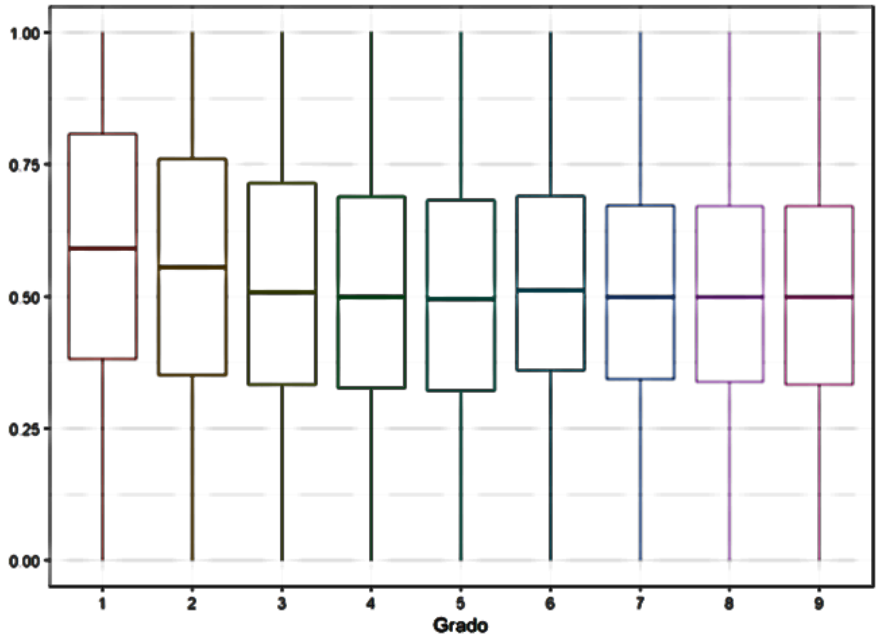
Estadístico	Irrelevante	Pequeño	Medio	Grande
Eta parcial	<0,01	0,01	0,06	0,14

Fuente: Tomado y adaptado de Goss & Meneses (2018),

Resultados

Como primer acercamiento, se presentan los descriptivos sobre la distribución de los grados, de primero a noveno, en la variable evaluación escalada, que define las notas de los estudiantes en una escala de cero (0) a (1) (Figura 13). Acá se evidencia una leve tendencia a la baja en el cuarto grado respecto a los grados iniciales donde los datos son medio altos, mientras que de cuarto grado en adelante tienen a tomar valores muy cercanos al promedio. Por otra parte, la distribución de los datos en cada grado es ligeramente distinta entre los grados iniciales, donde hay mayor presencia de datos extremos, mientras que de sexto en adelante las distribuciones presentan mayor agrupamiento sin ser una diferencia muy grande.

Figura 13. Distribución de los grados en la variable escalada evaluación en Lenguaje.



Fuente: Elaboración propia a partir de los datos del SAE.

Como se mencionó con anterioridad, una de las características de los datos del SAE es que cada estudiante no tiene una cadena de registros completa desde el grado primero

hasta el grado noveno, razón por la que hay una cantidad considerable de datos perdidos. La tabla 5 informa sobre los grupos con mayor cantidad de datos con los cuales se realizó la investigación.

Tabla 5. Grupos con mayor cantidad de datos para la realización de Anova

Grupo	Agrupaciones de grados	Número de estudiantes de cada grupo
1	"3", "4", "5"	19.291
2	"2", "3", "4"	8067
3	"1", "2", "3", "4"	8002
4	"2", "3", "4", "5"	7899
5	"3", "4", "5", "6"	2461
6	"1", "2", "3", "4", "5"	1416
7	"2", "3", "4", "5", "6"	1323

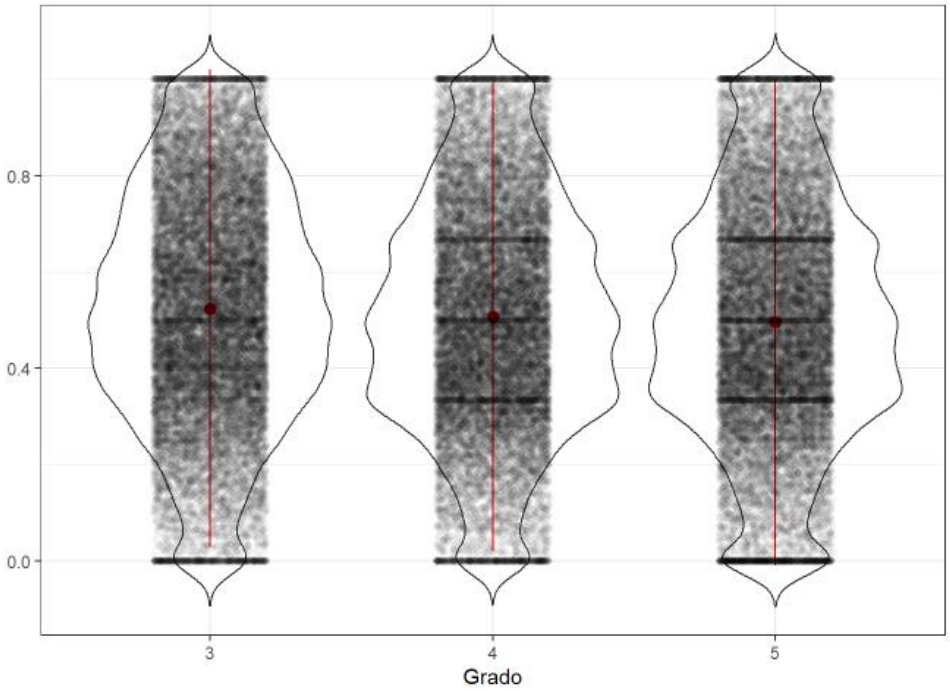
Fuente: Elaboración propia a partir de los datos del SAE.

Respecto al grupo 1, definido por los estudiantes que tienen registros para los grados 3, 4 y 5, la prueba de esfericidad de Mauchly indica que se ha cumplido el supuesto de esfericidad, MW= 0,94; p valor= 4.7e-66 (>0,05). Los resultados del Anova indican que existen diferencias estadísticamente significativas (<0,05), particularmente hay diferencias entre los pares, (3 y 4; p valor= 0.0000000000000002) (3 y 5; p valor= 0.0000000000000002) y (5 y 4; p valor= 0.000000001).

La Figura 14 describe la distribución de estos tres cursos (3. 4 y 5). En este se puede notar que los datos son más cercanos a la media en el grado tercero mientras que para el

grado cuarto y quinto el primer y cuarto cuartil toma valores más hacía los extremos, respecto a la media, aun cuando el Anova indica una diferencia, gráficamente no es muy grande, lo que si resulta distinta es la distribución tendiente a la asimetría en el tercer grado y con mayor variabilidad para el cuarto y quinto grado.

Figura 14. Distribución de la agrupación del grupo 1 para los grados 3, 4 y 5

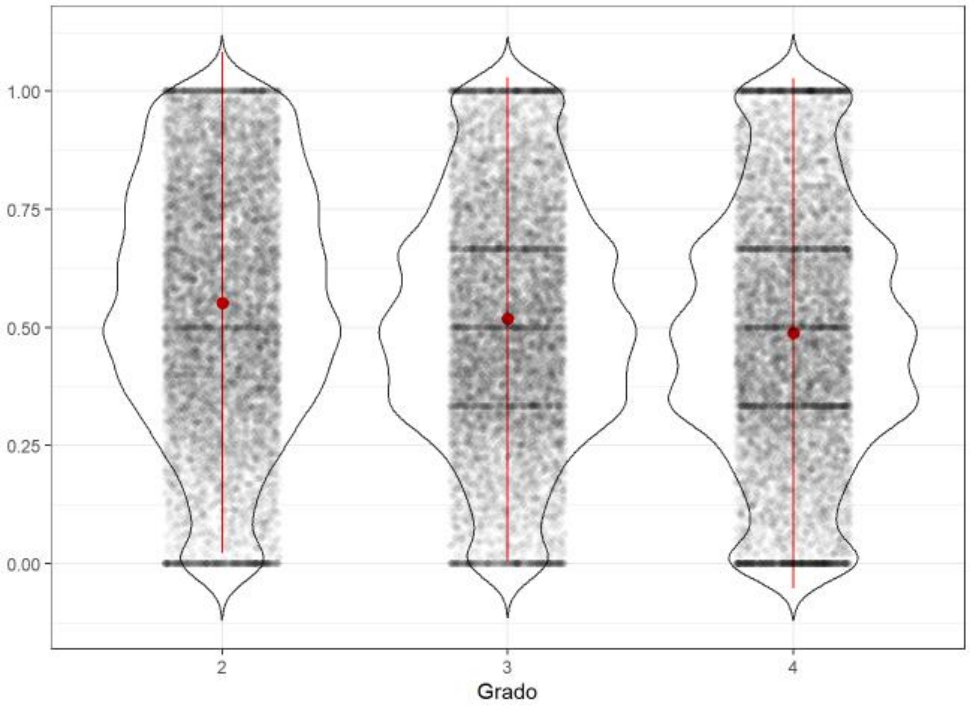


Fuente: Elaboración propia a partir de los datos del SAE.

Respecto al grupo 2, definido por los estudiantes que tienen registros para los grados 2, 3 y 4, la prueba de esfericidad de Mauchly indica que se cumple el supuesto de esfericidad, MW= 0,96; p valor= >0,05. Por lo tanto, el resultado del Anova indica que existen diferencias estadísticamente significativas (<0,05), particularmente hay diferencias significativas entre los pares, (2 y 3; p valor= 0.0000000000000002) (2 y 4; p valor= 0.0000000000000002) y (3 y 4; p valor= 0.0000000000002).

La Figura 15 describe la distribución de estos tres cursos (2. 3 y 4). En este se puede notar una distribución similar a la figura anterior, sin embargo, en esta figura es más visible la tendencia a la baja desde el segundo al cuarto grado. También resulta de interés cómo los primeros grados son más simétricos y presentan menos ondulaciones en cuando a la distribución. Otro punto diferencial entre segundo versus tercero y cuarto es la que distribución de los datos en los primeros grados tiende a estar por encima de la media y la presencia de datos medio-bajos o bajos es menos frecuente, tendencia cambiante con el pasar de los grados.

Figura 15. Distribución de la agrupación del grupo 2 para los grados 2, 3 y 4.

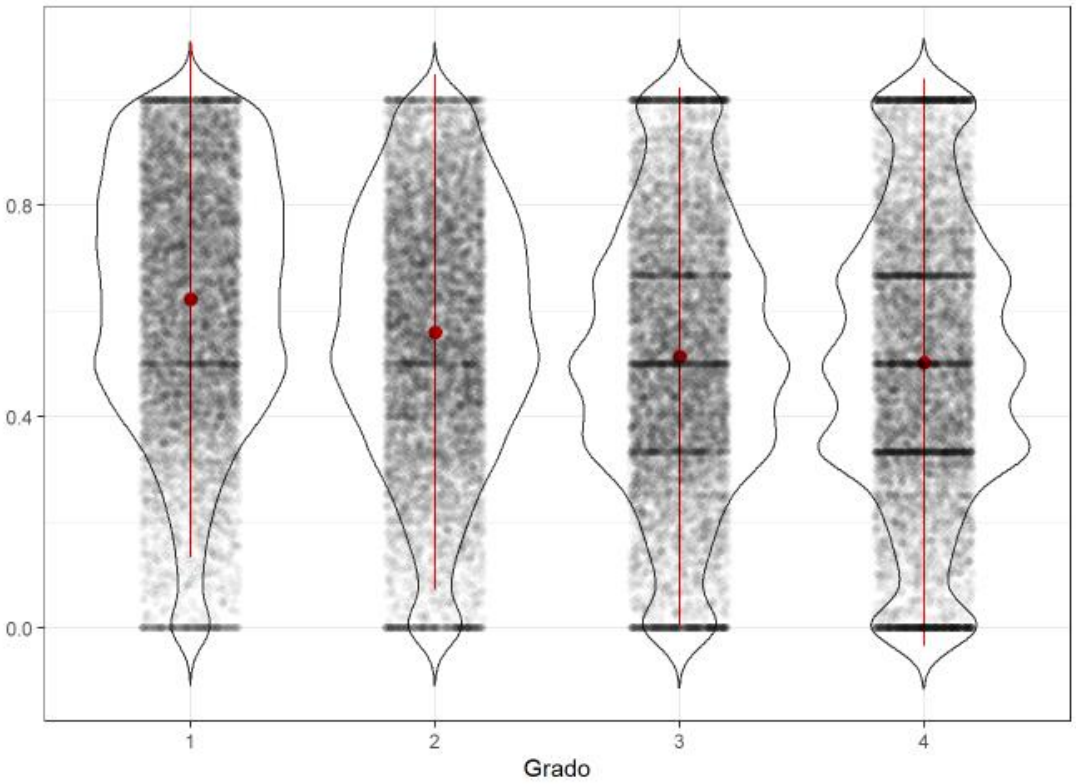


Fuente: Elaboración propia a partir de los datos del SAE.

Respecto al grupo 3, definido por los estudiantes que tienen registros para los grados 1, 2, 3 y 4, la prueba de esfericidad de Mauchly indica que se cumple el supuesto de esfericidad, MW= 0.898; p valor= 9.86e-185 (>0,05), Por lo tanto, el resultado del Anova

indica que existen diferencias estadísticamente significativas (<0,05), particularmente hay diferencias entre todos los pares.

Figura 16. Distribución de la agrupación del grupo 3 para los grados 1, 2, 3 y 4.

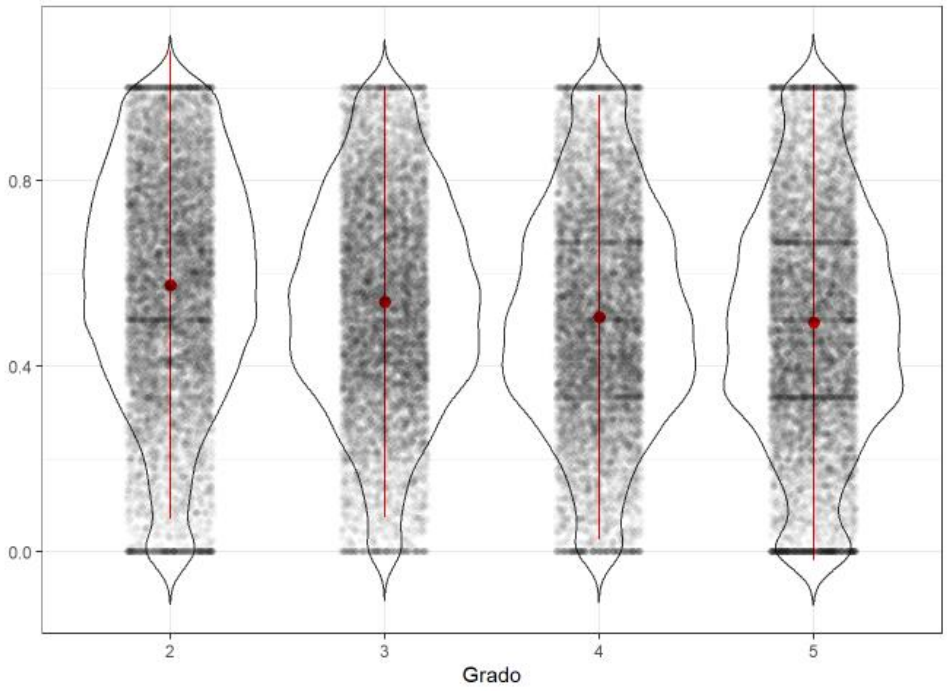


Fuente: Elaboración propia a partir de los datos del SAE.

La Figura 16 describe la distribución de estos cuatro cursos (1, 2. 3 y 4). Se evidencia de manera más prominente la tendencia a la baja, muy marcada si se compara el primer grado escolar y el cuarto grado escolar, de igual modo, también resulta coherente con la distribución asimétrica en los grados iniciales y su desproporción con el avance de los grados. Otro punto de interés es el comportamiento de los valores extremos, particularmente los que representan las bajas calificaciones, que suelen ser más frecuentes a medida que aumentan los grados escolares.

Respecto al grupo 4, definido por los estudiantes que tienen registros para los grados 2, 3, 4 y 5, la prueba de esfericidad de Mauchly indica que se cumple el supuesto de esfericidad, MW= 0.932; p valor= 1.45e-117 (>0,05), Por lo tanto, el resultado del Anova indica que existen diferencias estadísticamente significativas (<0,05), particularmente hay diferencias entre todos los pares.

Figura 17. Distribución de la agrupación del grupo 4 para los grados 2, 3, 4 y 5.



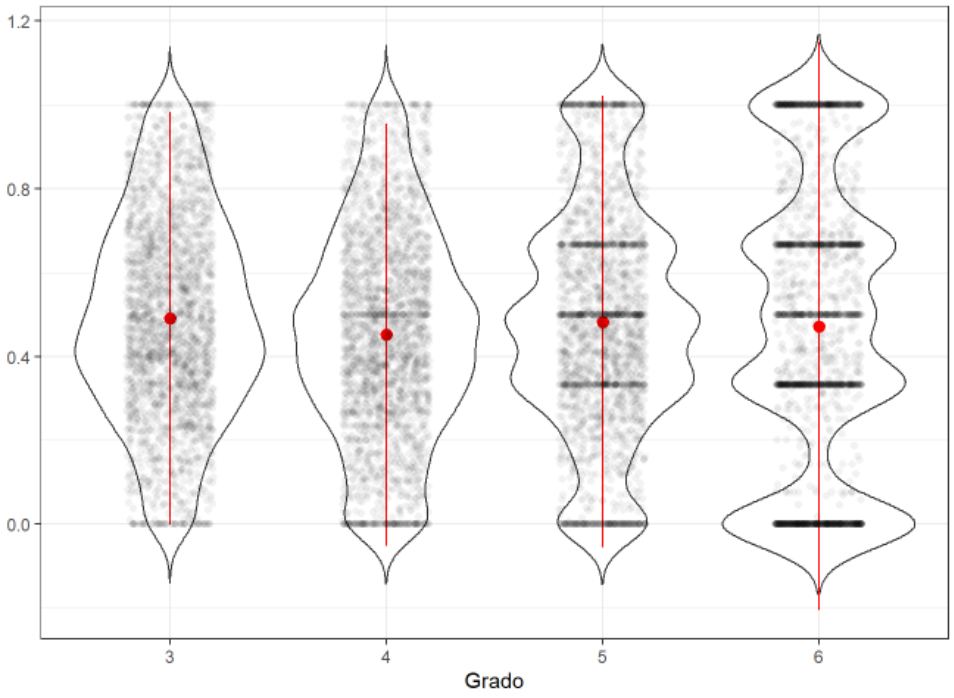
Fuente: Elaboración propia a partir de los datos del SAE.

La Figura 17 describe la distribución de estos cuatro cursos (2, 3, 4 y 5). Esta agrupación en particular resulta la más estable respecto a la simetría, sin embargo, es evidente la tendencia a la baja, es más notorio el cómo las evaluaciones de los primeros grados tienden a agruparse con mayor frecuencia en los datos medio altos teniendo cuartil uno y cuartil 3 por encima del valor medio (0,5), tendencia que decrece con el aumentar de

los grados escolares. También se puede notar que los valores extremos inferiores son más frecuentes con el pasar de los cursos.

Respecto al grupo 5, definido por los estudiantes que tienen registros para los grados 3, 4, 5 y 6 la prueba de esfericidad de Mauchly indica que se cumple el supuesto de esfericidad, MW= 0.773; p valor= 1.66e-134 (>0,05), Por lo tanto, el resultado del Anova indica que existen diferencias estadísticamente significativas (<0,05), particularmente hay diferencias entre los pares (3 y 4) y (4 y 5), en los otros pares no son significativas las diferencias.

Figura 18. Distribución de la agrupación del grupo 5 para los grados 3, 4, 5 y 6.



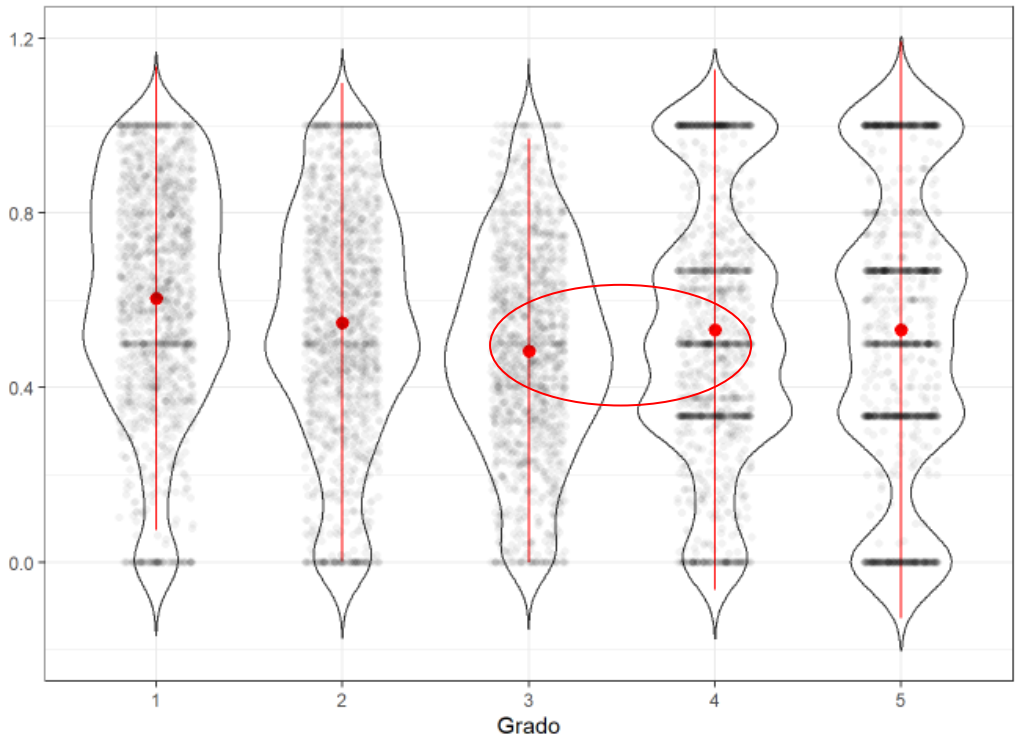
Fuente: Elaboración propia a partir de los datos del SAE.

La Figura 18 describe la distribución de estos cuatro cursos (3, 4, 5 y 6). Esta agrupación en particular muestra un dato de secundaría, que no se había registrado en las anteriores agrupaciones. Se aprecia que tercero y cuarto muestran una tendencia similar,

mientras que quinto y sexto muestran una distribución distinta con mayor variabilidad. Adicional a lo encontrado en las anteriores agrupaciones, en esta figura es más notable como las notas en los grados mayores presentan información de alta frecuencia en los valores extremos, ahora tanto en la parte superior como inferior. También se encuentra que el cuarto grado es el de menor promedio, y en grado quinto y sexto aumenta y se estabiliza.

Respecto al grupo 6, definido por los estudiantes que tienen registros para los grados 1, 2, 3, 4, 5, la prueba de esfericidad de Mauchly indica que se cumple el supuesto de esfericidad, MW= 0.68; p valor= 3.85e-109 (>0,05). Por lo tanto, el resultado del ANOVA indica que existen diferencias estadísticamente significativas (<0,05), particularmente hay diferencias entre todos los pares, excepto en el par (4 y 5) p valor= 0,9.

Figura 19. Distribución de la agrupación del grupo 6 para los grados 1, 2, 3, 4 y 5.



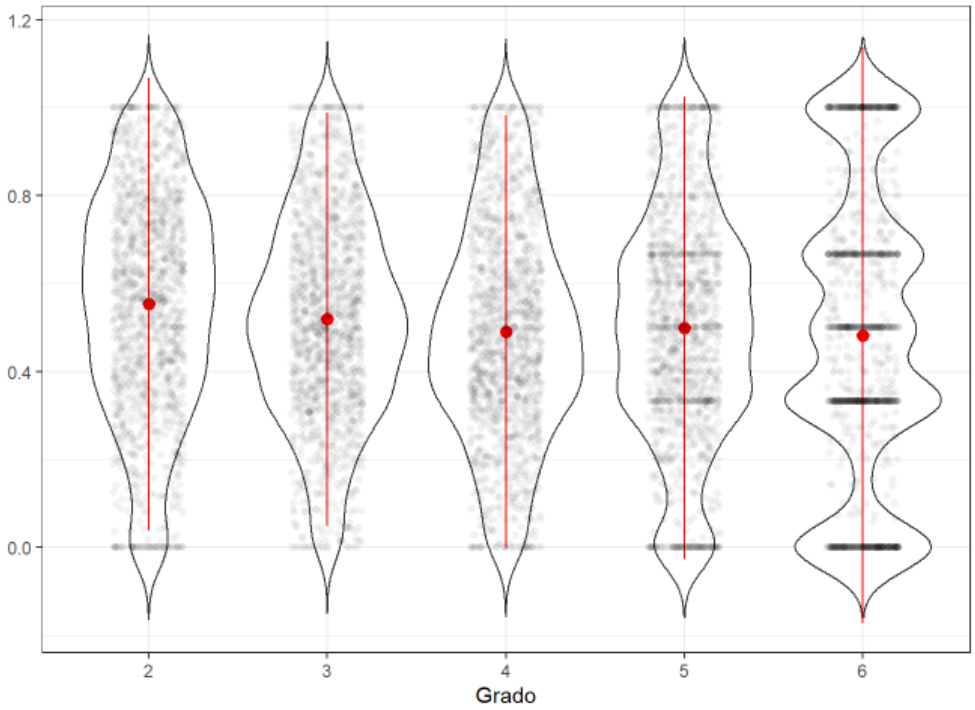
Fuente: Elaboración propia a partir de los datos del SAE.

La Figura 19 describe la distribución de estos cinco cursos (1, 2, 3, 4 y 5). Esta agrupación resulta consistente con lo ya reportado respecto a la distribución de los datos, el comportamiento de los valores extremos, sin embargo, la tendencia a la baja se da en tercero y no en el cuarto grado escolar. Se distinguen dos patrones, de primero a tercero, una distribución de datos con menores latencias, y otro patrón de cuarto a quinto, con mayores latencias en la distribución, pero con una media estable en los dos grados escolares.

Finalmente, respecto al grupo 7, definido por los estudiantes que tienen registros para los grados 2, 3, 4, 5 y 6 la prueba de esfericidad de Mauchly indica que se cumple el supuesto de esfericidad, MW= 0.657; p valor= 8.96e-114 (>0,05). Por lo tanto, el resultado del Anova indica que existen diferencias estadísticamente significativas (<0,05), particularmente no hay diferencias entre los pares (3 y 5), (4 y 5), (4 y 6), y (5 y 6), en las otras combinaciones si hay diferencias estadísticamente significativas.

La Figura 20 describe la distribución de estos cinco cursos (2, 3, 4, 5 y 6). Esta agrupación además de ser consistente con las agrupaciones anteriores permite ver una diferencia en la distribución entre los grados de primaría y el grado de segundaría, donde hay unos puntos con una frecuencia importante que, en la lógica de la escala de cero a cinco para ejemplificar, podría representar una alta frecuencia de la nota cinco (5), un valor cercano al dos (2), y el cero (0).

Figura 20. Distribución de la agrupación del grupo 7 para los grados 2, 3, 4, 5 y 6.



Fuente: Elaboración propia a partir de los datos del SAE.

En términos generales, las siete agrupaciones muestran resultados consistentes entre sí, en general se encontraron diferencias estadísticamente significativas en todos los grupos, y a nivel de pruebas *posthoc* también se encontraron estas diferencias entre los pares comparados, a excepción de los grados sextos frente a grados más recientes como quinto o cuarto, pero siempre hubo diferencias entre los grados iniciales y los grados finales, en particular sobre el cuarto grado que resulta el interés de esta investigación.

Respecto al análisis gráfico, las siete agrupaciones muestran resultados consistentes en varios aspectos; un primer aspecto es la tendencia a la baja entre los primeros grados hasta el cuarto o quinto grado escolar, en todas las agrupaciones la pendiente que se pueda trazar es negativa, que se tiende a estabilizar en los últimos grados registrados, quinto y sexto. Un segundo aspecto interesante es como se agrupan los datos con mayor frecuencia

hacia la parte promedio-alto en los grados iniciales y como esta tendencia va disminuyendo grado tras grado. Un tercer elemento está relacionado con los valores extremos, en particular con el extremo inferior para primaria, pero equilibrado entre extremo superior en inferior para el sexto grado, tendiendo mayor prevalencia con el aumento de los grados puntuaciones altas, como cinco (5), o cien (100) en esas escalas, y puntuaciones inferiores como el cero (0). Un cuarto elemento son los patrones en la distribución, que en algunos casos conformaba los grados iniciales hasta el tercer grado frente a otro patrón de distribución para el cuarto o quinto grado, particularmente el quinto. También podría notarse la diferencia entre la primaria y la segundaria en la distribución de la información.

Discusión y conclusiones

Los resultados de la investigación permiten afirmar que se presenta el Declive de Cuarto Grado en la población estudiada. Esta información resulta coherente con el informe de resultados de PIRS 2011 realizado por el ICFES para los datos colombianos donde se comenta que el cuarto grado es un periodo de tránsito determinante en el desarrollo lector de los estudiantes, particularmente porque implica un paso de aprender a leer, en los grados primero, segundo y tercero, para a partir de cuarto grado, utilizar la lectura como un medio para acceder al conocimiento, agrupaciones que también se encontraron en los resultados de este estudio.

De igual forma, la pendiente negativa o tendencia a la baja desde grado primero hasta cuarto o quinto grado, corresponde a lo que inicialmente Walter Loban (1956) identificaba entre los años 50´s y 60´s, donde ya se hablaba del cuarto grado como un grado *embudo* donde las brechas en las habilidades de lectura se harían más notorias y sería un punto de inicio a una brecha mucho mayor. La investigación de Loban (1956) describió información hasta el grado 12, dando descripción de cómo se comportaba esa brecha en la segundaria respecto a la primaria, sin embargo, esta investigación soló brindó información hasta sexto grado para los Anova por características propias del diseño de investigación y las herramientas de análisis propuestas.

También es consistente con el estudio de mayor impacto y de amplio conocimiento en el área como lo es la investigación realizada por Chall & Jacops (1983), quienes fueron los precursores del concepto de *Fourth Grade Slump*, de ello pueden relacionarse la distribución de las notas, donde la tendencia en los grados iniciales es el promedio-alto, y a medida que

se acerca el cuarto grado, y luego de este, la tendencia de las notas era hacia un promedio bajo.

Con estos resultados, cobra sentido lo mencionado por (Chall & Jacops, 1983), respecto a que estas primeras etapas se caracterizan por la prelectura, centrada en la codificación, textos sencillos y familiares con una coherencia sencilla y una alta cohesión, mientras que el segundo patrón lleva al estudiante a enfrentarse a temas más complejos, lingüísticamente y cognitivamente más demandantes, con nuevas palabras que exigen un mayor nivel de análisis y de vocabulario, pero lo más relevante es que el propósito de la lectura cambia, para los grados cuarto en adelante la lectura se convierte en un medio para acceder a un conocimiento, la puerta para el aprendizaje del mundo de conocimiento.

De igual forma, las tendencias a que los valores centrales de distribución se concentren en valores medio-bajos también es diciente de bajos resultados en la prueba SABER para tercero y cuarto, donde la mitad de los estudiantes se encuentran en los niveles insuficientes y mínimos. (ICFES, 2018). Del mismo modo para las pruebas internacionales (PISA), donde Colombia ha estado por debajo del promedio de los países miembros, y también para otra prueba internacional como PIRLS, donde solo en 1% de los estudiantes de Colombia alcanza un nivel avanzado en comprensión lectora y el 28% no alcanza los puntajes mínimos (PIRLS. 2011)

Por otro lado, las gráficas descritas en este estudio permiten ver que, al trazar una recta en las diferentes agrupaciones para todos los grados, el cuarto grado fue un punto de inflexión, información que resulta coherente cuando se describe que este grado, según datos del DANE (2012), es el grado en el que más se presenta disminución de matrículas, por tanto, podría hipotetizarse que implica una carga tan alta que se relaciona con la deserción escolar.

Respecto a otro punto, que menciona que el declive es solo un punto de inflexión del cual se recuperan los estudiantes, como lo describe Etmanskie & Cols (2009), quienes indican que los estudios sobrestiman el bajón en el rendimiento, pues luego de esta etapa de tránsito los estudiantes obtienen nuevamente resultados adecuados. Frente a ello los datos de esta investigación muestran una tendencia a la baja, información que cobra relevancia cuando la fuente de información es institucional y masiva. Claramente la diferencia no es de dimensiones mayores, como lo estima el Tamaño del efecto, que resultó pequeño o insignificante en la mayoría de las agrupaciones, pero es una diferencia importante que marca una tendencia a la baja que parece mantenerse constante, pero sobre el promedio bajo.

Relacionado con lo anterior, otros estudios, como el de Spindler, Jørgensen & Holmen, (2014) señalan que el declive no es una cuestión individual sino producto de las actividades escolares que se plantean dentro de la institución, pues están no permiten beneficiar al estudiante para superar esta etapa de tránsito. De igual modo, Donaldson, (2010), encontró en su investigación que, el hecho que los estudiantes participaran en un programa estatal, Success for All program, implicaba encontrar una menor incidencia del declive,

En particular, los resultados de esta investigación han de incidir en la promoción y dirección de la política pública, en un primer nivel frente a las decisiones que pueda tomar la Secretaría de Educación de Bogotá como entidad rectora, que, de la mano del Ministerio de Educación, definen y trazan la política pública de Bogotá y de Colombia. Por ejemplo, los proyectos para hacer evaluación masiva en el cuarto grado (*Evaluar para avanzar*) o algunas estrategias desarrolladas por el ICFES como entidad encargada de la evaluación en

Colombia, ha de ser orientada en hacer evaluación en el cuarto grado como lo hace PIRSL y otras evaluaciones internacionales, también es importante orientar el declive como modelo explicativo dentro de estrategias de alcance nacional como estrategias del programa Colombia aprende, programa nacional de lectura y escritura "leer es mi cuento", entre otras estrategias de alto impacto nacional.

A nivel de centro, esta información es de especial relevancia para el docente, pues le va permitir entender que el cuarto grado es una etapa diferencial y que sus recursos pedagógicos han de orientarse a la adaptación y al afrontamiento de esta etapa, ya que el primer paso para establecer un plan de mejoramiento es el diagnóstico y la descripción del problema, y si no se tiene conocimiento de que el Declive de Cuarto Grado es un fenómeno presente en los estudiantes, va a ser difícil que se tomen acciones frente al mismo. (Etmanskie & Cols,2009)

De igual modo para el estudiante y para sus familiares es una información de especial relevancia, dado que ayuda a comprender el porqué del bajo rendimiento de algunos de ellos, por qué se le facilitan los contenidos narrativos, o los familiares, de alta cohesión, de vocabulario de uso frecuente, entre otras características, y por qué resulta retador enfrentarse a contenidos como los textos argumentativos, expositivos, con palabras de baja frecuencia y textos de baja cohesión. (Hirsch, 2003) Todo ello hace parte de lo que implica el declive de cuarto grado, pero no es reconocido como tal desde el nivel institucional, ni desde el nivel docente, ni del nivel estudiante ni desde el nivel familiar, y es un aspecto de especial trascendencia cuando del rendimiento escolar se trata.

En general esta investigación toma valor por la representatividad y alcance de sus conclusiones ya que toma como partida que el elemento a analizar es una información bruta registrada por los docentes de la ciudad de Bogotá mediante un recurso institucional, el SAE. Por tanto, el número de registros tan alto, a pesar de las depuraciones realizadas, da una mayor fuerza y garantía de validez a las conclusiones y a los resultados presentados.

Sin embargo, también es importante mencionar las limitaciones del estudio y las proyecciones frente al desarrollo de este. Una de las mayores limitaciones a nivel operativo y a nivel de validez de lo que se concluye es que, por tratarse de un estudio exposfacto, la recolección de datos se realiza luego de que el suceso haya ocurrido, en este caso, la matriz de datos del SAE no es un matriz que se haya generado para realizar esta investigación, como se mencionó, es un insumo institucional que ha tenido durante su vigencia poco espacio para la investigación, siendo esta investigación una de las pioneras. Ello llevo a que fuese necesario hacer un ejercicio de depuración y varios procedimientos de transformación los datos para su análisis, hasta contar con el escenario que fuese más representativo y fiel a la información presentada, aspecto que resultó complejo respecto a los recursos que demandó, no obstante, la trasformación min-max y todo el procedimiento generado se realizaron con toda la calidad metodológica correspondiente.

Relacionado con ello, una limitante importante es que no se haya podido dar información de lo que sucede en sexto grado en adelante más allá de los datos descriptivos, pues aun cuando los datos del SAE recolectan información desde primer grado hasta noveno grado, aspecto que conecta directamente con la proyección del estudio, ya que tener la cadena de registros completa permite aumentar la potencia y validez de los datos y

desarrollar modelos estadísticos más robustos, como un análisis de series de tiempo por ejemplo, que describiría con más detalle el fenómeno de interés.

Otro aspecto a considerar dentro de las limitaciones del estudio corresponde al sesgo derivado de tomar las calificaciones de los estudiantes en las asignaturas relacionadas con "lengua castellana" como variable que da cuenta del rendimiento en la comprensión de lectura y a partir de ello hacer inferencias sobre el declive, ello en razón de que las calificaciones muestran parcialmente la habilidad para alcanzar unos criterios y no necesariamente un aprendizaje de los estudiantes, pues las variables mediadoras son diversas y complejas. No obstante, partiendo de la representatividad de los datos y de los resultados estadísticos encontrados, esta investigación si podría indicar alguna tendencia que resulta como indicador del declive en estas habilidades necesarias para el aprendizaje viéndose reflejadas en un declive en las notas, particularmente en lo relacionado con "lengua castellana".

También es importante que futuros estudios indaguen, ya sea mediante encuestas, entrevistas o por otros medios, por cómo los docentes reconocen este fenómeno y cómo actúan sobre él, porque lo más seguro es que sea algo evidente desde la experiencia, pero no un fenómeno conocido desde la academia. Futuros estudios deben indagar por cómo se ha comprendido este bajo rendimiento en el cierre de la etapa de primaria y como ella repercute en el futuro rendimiento en segundaría, tal vez extendiéndolo hasta la educación universitaria, en la medida que las habilidades de comprensión de lectura influyen en muchas variables dentro del ciclo de desarrollo. (Gómez & Cols, 2011)

Finalmente, otra proyección de este estudio son las estrategias de intervención y los planes de mejora que se pueden plantear una vez se reconozca la presencia del fenómeno. Y frente a ello la investigación especializada en el área tiene importantes desarrollos, (Dnaldson,2010), (Delsing, 2020), (Etmanskie. 2009). pues como se mencionó con anterioridad, la investigación sobre el declive se hace centrado en estos últimos tiempos en desarrollar intervenciones que en demostrar su prevalencia.

Implicaciones éticas

De acuerdo a la ley 1090 del 2006, por la cual se reglamenta el ejercicio de la profesión de Psicología, se dicta el Código Deontológico y Bioético y otras disposiciones, este ejercicio académico mantendrá los más altos estándares profesionales, brindando información completa y suficiente a los participantes del estudio, garantizando la confidencialidad de la información respetando el derecho de los usuarios de conocer los resultados, las interpretaciones hechas y las bases de sus conclusiones y recomendaciones. Asegurando la seguridad de la información recolectada y de otras técnicas de evaluación dentro de los límites de los mandatos legales. Finalmente, es pertinente resaltar que se contó con la aprobación de la entidad encargada de los datos, la Secretaría de Educación de Bogotá, firmando los acuerdos de confidencialidad y asegurando el manejo correspondiente legalmente de la información por parte del investigador, de igual modos, los datos fueron anonimizados directamente por la entidad y usados por el investigador con fines estrictamente académicos.

Referencias

Achenbach, T. M., & Rescorla, L. A. (2006). Developmental issues in assessment, taxonomy, and diagnosis of psychopathology: Life span and multicultural perspectives.

Arnau, J., & Bono, R. (2008). Estudios longitudinales de medidas repetidas: Modelos de diseño y análisis. Escritos de Psicología (Internet), 2(1), 32-41.

Asociación Internacional para la Evaluación del Logro-Educativo (IEA), (2022). PIRLS. Progreso en el Estudio Internacional de Competencia Lectora, consultado 4 de junio del 2022, url: https://www.iea.nl/studies/iea/pirls

Best, R., Floyd, R. G., & McNamara, D. S. (2004). Understanding the fourth-grade slump: Comprehension difficulties as a function of reader aptitudes and text genre. *In 85th Annual Meeting of the American Educational Research Association.*

Bologna, E. (2013). Estadística para psicología y educación. Editorial Brujas.

Bogaerds-Hazenberg, S. T., Evers-Vermeul, J., & van den Bergh, H. (2020). A Meta-Analysis on the Effects of Text Structure Instruction on Reading Comprehension in the Upper Elementary Grades. *Reading Research Quarterly*.

Balluerka, N. & Vergara, A. (2002). Diseños de investigación experimental en psicología. Madrid: *Prentice Hall.*

Canal, Neus. (SF) Distribuciones de probabilidad. El teorema central del límite. Capítulo 8. Disponible en: https://www.revistaseden.org/files/8-CAP%208.pdf

Chall, J.S., & Jacobs, V.A. (1983). Writing and reading in the elementary grades: Developmental trends among low-SES children. *Language Arts*, vol. 60, no. 5, pp. 617-626.

Chall, J. S., & Jacobs, V. A. (2003). The classic study on poor children's fourth-grade slump. *American educator*, *27*(1), 14-15.

Castillo, R., & Bravo, A. A. (2015). Importancia del tamaño del efecto en el análisis de datos de investigación en psicología. Persona: Revista de la facultad de psicología, (18), 137-148.

Dehaene, S. (2019). Aprender a leer: de las ciencias cognitivas al aula. Siglo XXI Editores.

Delsing, Maria, "From Fourth Grade Slump to Fourth Grade Triumph" (2010). Ed.D. Dissertations. 3. https://digitalcommons.olivet.edu/edd_diss/3

Donaldson, N. M. (2010). The fourth grade slump: The relationship between reading attitudes and frequency of reading (*Doctoral dissertation, Bowling Green State University*).

Etmanskie, J. M (2009). A longitudinal examination of the persistence of late emerging reading disabilities. *Journal of learning disabilities*, Tesis DOCTOR OF PHILOSOPHY*. THE UNIVERSITY OF BRITISH COLUMBIA (Vancouver) April 2009*

Etmanskie, J. M., Partanen, M., & Siegel, L. S. (2016). A longitudinal examination of the persistence of late emerging reading disabilities. *Journal of learning disabilities*, *49*(1), 21-35.

García, L. O. (2016). Nivel socioeconómico y rendimiento académico: estudiantes resilientes (Doctoral dissertation, Universidad Complutense de Madrid).

Gómez, S. M. C., Gutiérrez, D. M. G., & Hassan, A. R. (2011). Determinantes del rendimiento académico en Colombia: pruebas ICFES Saber 11º, 20091. *Recuperado de*

http://www. eafit. edu. co/escuelas/economiay finanzas/publicaciones/Documents/workingpapers/Determinantes% 20del% 20rendimiento% 20acad% C3% A9mico% 20en% 20Colombia% 20pruebas% 20ICFES% 20Saber, C2.

Goss-Sampson, M. A., & Meneses, J. (2018). Análisis estadístico con JASP: Una guía para estudiantes.

Gordillo, R. C., Mayo, N. C., Lara, G. G., & Gigante, S. V. (2010). Metodología de la investigación educativa: Investigación ex post facto. Barcelo: https://www. uam. es/personal_pdi/stmaria/jmurillo/InvestigacionEE/Presentaciones/Curso_10/EX-POST-FACTO_Trabajo. Pdf.

Hecke, T. V. (2012). Power study of anova versus Kruskal-Wallis test. *Journal of Statistics and Management Systems,* 15(2-3), 241-247.

Hirsch, E. D. (2007). La comprensión lectora requiere conocimiento de vocabulario y del mundo. *Estudios públicos*, (108).

Hirsch, E. D. (2003). Reading comprehension requires knowledge of words and the world. *American Educator*, *27*(1), 10-13.

Instituto Colombiano para la Evaluación de la Educación (ICFES) (2018) Informe de resultados nacionales Saber 3°, 5° y 9° 2012 – 2017. 978-958-11-0783-4 Bogotá D.C., abril de 2018.

Janus Spindler Møller, J. Normann Jørgensen & Anne Holmen (2014) Polylingual development among Turkish speakers in a Danish primary school – a critical view on the fourth grade slump, *International Journal of Bilingual Education and*

Bilingualism, 17:1, 32-54,

DOI: 10.1080/13670050.2012.735221: https://doi.org/10.1080/13670050.2012.735221

Organización para la Cooperación y el Desarrollo Económico (OCDE) (2018) Resultados 2018 Programa para la Evaluación Internacional de Alumnos (PISA) Notas para Colombia.

Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. Journal of educational psychology, 95(2), 211.

Ledesma, R., Macbeth, G., & CORTADA DE KOHAN, N. U. R. I. A. (2008). Tamaño del efecto: revisión teórica y aplicaciones con el sistema estadístico ViSta. *Revista latinoamericana de psicología, 40(3), 425-439.*

Llorca, Javier, & Delgado, Miguel (2004). Estudios longitudinales: concepto y particularidades. Revista Española de Salud Pública, 78(2),141-148. [fecha de Consulta 22 de mayo de 2022]. ISSN: 1135-5727. Disponible en: https://www.redalyc.org/articulo.oa?id=17078202

Lori Helman, Anne Delbridge, David Parker, Martina Arnal & Luz Jara Mödinger (2016) Measuring Spanish orthographic development in private, public and subsidised schools in Chile, *Assessment in Education: Principles, Policy & Practice,* 23:3, 327-352, DOI: 10.1080/0969594X.2015.1038217 https://doi.org/10.1080/0969594X.2015.1038217

Martín-Lagos López, M. D. (2018). Educación y desigualdad: una metasíntesis tras el 50 aniversario del Informe Coleman. *Revista de Educación, 380. Abril-junio 2018, pp. 186-209.*

Martín Martínez, S. G. (2015). Pruebas Saber de lenguaje 3° y 5°: posibilidades y retos desde la perspectiva de la evaluación formativa. *Tesis de grado Magister en Educación con Énfasis en Evaluación y Gestión Educativa*.

Marqués, I. (2016). Apuntes sobre el Informe Coleman. Sobre la difícil convivencia de los principios igualitarios en un Mundo Desigual*. International Journal of Sociology of Education,* 5(2), 107-126.

Marin, J. M., & ANOVA, U. (2019). Capítulo 16, Análisis de varianza con medidas repetidas. El procedimiento modelo lineal general: medidas repetidas. [online] http://halweb.uc3m.es/esp/Personal/personas/ jmmarin/esp/GuiaSPSS/16anovar.pdf

McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly*, 30(1), 934-956.

Muñoz Valenzuela, Carla, Ferreira Torres, Schonemann, Sánchez Quintul, Pamela, Santander Pérez, Sylvia, Pérez Rodríguez, Marcela, & Valenzuela Carreño, Jorge. (2012). Características psicométricas de una escala para caracterizar la motivación por la lectura académica. *Revista electrónica de investigación educativa*, 14(2), 118-132.

McNamara, D., Ozuru, Y., & Floyd, R. (2017). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education, 4(1), 229-257.*

Peñaloza, H. A. B., & Alba, A. L. (2014). La influencia de las TIC en el desempeño académico: evidencia de la prueba PIRLS en Colombia 2011. *Revista academia y virtualidad,* 7(2), 15-26.

Rodríguez, L. J. C. (2020). Efectos del coronavirus en el sistema de enseñanza: aumenta la desigualdad de oportunidades educativas en España. *Revista de Sociología de la Educación-RASE*, *13*(2), 114-139.

Román, M., & Murillo, F. J. (2014). Disponibilidad y uso de TIC en escuelas latinoamericanas: incidencia en el rendimiento escolar. Educação e Pesquisa, 40(4), 879-895.

Secretaría de Educación de Bogotá (2022). *Presentación_SAE_Investigación* [Diapositivas de PowerPoint]. Dirección de Evaluación Secretaría de Educación de Bogotá,

Suhr, K. A., Hernandez, D. A., Grimes, D., & Warschauer, M. (2010). Laptops and fourth grade literacy: Assisting the jump over the fourth-grade slump. *The Journal of Technology, Learning and Assessment, 9(5).*

Pardo, A.; Ruiz M.A. y San Martín, R. (2015). Análisis de datos en ciencias sociales y de la salud (vol I, 2ª ed.). Madrid: Síntesis. Pardo, A y San Martín, R (2015). Análisis de datos en ciencias sociales y de la salud (vol II, 2ª ed). Madrid: Síntesis.