

**PRUEBA MECÁNICA
DE LECTURA:
ANÁLISIS DE
VALIDACIÓN Y
FIABILIDAD**

Diciembre 2023

Presentado a:

FUNDACIÓN FLAMBOYÁN

Presentado por:

JULIÁN LABOY RODRÍGUEZ

✉ jlaboy@datosenperspectiva.com

JESÚS MARRERO CENTENO

✉ jmarrero@datosenperspectiva.com

TABLA DE CONTENIDO

Introducción	6
Metodología	7
Análisis descriptivo y resultados	9
Análisis de validez y fiabilidad	23
Resumen de hallazgos	34
Conclusión	36
Referencias	39

RESUMEN EJECUTIVO

Se administró una prueba de mecánica de lectura a estudiantes de grados primarios (1ro a 3ro) en dos momentos diferentes, febrero de 2023 y mayo de 2023. Se realizaron esfuerzos exhaustivos para analizar los datos de validez y confiabilidad de la prueba. Los resultados obtenidos fueron muy prometedores, demostrando no solamente la utilidad de la prueba por su consistencia interna para medir dicha destreza, sino porque también evidenció mejoras significativas en las habilidades de lectura de los(las) estudiantes. A continuación, presentamos un resumen conciso de los principales hallazgos obtenidos de este estudio.

Hallazgos descriptivos

- Hubo aumentos en los promedios de mecánica de lectura de cada prueba al analizar el grupo en su totalidad (N = 30,477).
- Estos aumentos se sostuvieron al subdividir a los(las) estudiantes por grados (1ro a 3ro) y por regiones educativas.
- Encontramos evidencia que sugiere que ocurrió una mejora general en el rendimiento de los(las) estudiantes en sus destrezas de lectura (específicamente, la mecánica de lectura).

Aspectos de validez y confiabilidad

- **Equilibrio entre Consistencia y Mejora:**
 - Hubo una alta correlación entre ambos tiempos (febrero y mayo), lo cual muestra consistencia en los resultados.
 - Encontramos una mejora estadísticamente significativa en las puntuaciones entre ambos tiempos.
 - La magnitud de esas mejoras (tamaño de efecto) indica que hubo un cambio positivo en las habilidades de los estudiantes.
- **Relación con medidas similares, como calificaciones de Español**
 - A medida que mejoran las notas, los promedios de las subpruebas de mecánica de lectura también mejoraron.
 - Esto implica que hay una relación directa del contenido de la prueba con el contenido de la clase de Español.
 - Aunque la relación se encontró en todos los grados y regiones educativas, la misma fue mucho más directa para el primer grado.
- **Calidad de la Prueba y del Aprendizaje:**
 - La combinación de resultados sugiere que la prueba es una herramienta efectiva para medir la habilidad de lectura y que la instrucción proporcionada entre los dos momentos de la prueba ha sido beneficiosa para los estudiantes.
- **Evidencia de Aprendizaje:**
 - La combinación de un aumento en los promedios, la significancia estadística, los tamaños del efecto, así como la relación con las notas de Español, sugiere que es probable que haya habido aprendizaje o mejora en la habilidad de lectura entre los dos tiempos de prueba.

- **Relevancia Educativa:**

- Estos resultados pueden indicar que las lecciones o intervenciones educativas entre los dos tiempos de prueba fueron efectivas en mejorar las habilidades de lectura de los estudiantes.

El instrumento, por tanto, posee evidencia estadística y de contenido. Además, por su alineamiento con la materia de Español, sugerimos su continua administración. La calidad del instrumento evidencia que será de gran utilidad para que cada maestra(o) pueda saber el nivel específico en materia de lectura (y Español) de sus estudiantes y cómo intervenir de maneras informadas y apropiadas.

INTRODUCCIÓN

En el ámbito de la educación primaria, la evaluación precisa de las habilidades de lectura es fundamental para guiar el aprendizaje y el desarrollo de los estudiantes. En este reporte presentamos un análisis exhaustivo de la validación de una prueba de mecánica de lectura, diseñada para medir de manera confiable y efectiva estas habilidades esenciales. A través de una meticulosa revisión metodológica y estadística, exploramos los aspectos de confiabilidad y validez del instrumento, proporcionando una visión detallada de su capacidad para evaluar con precisión el progreso y las necesidades de los estudiantes en los primeros años de su trayectoria educativa.

En adelante, cubriremos aspectos clave como la metodología del estudio, los procesos de limpieza de datos y validación de la prueba, el análisis descriptivo y los resultados obtenidos. Profundizaremos en la evaluación de la consistencia y validez de la prueba, la correlación de las puntuaciones con las notas de español, las implicaciones educativas de los hallazgos, y ofreceremos unas conclusiones con recomendaciones y pasos futuros basados en los resultados del estudio.

METODOLOGÍA

Basamos este análisis en una metodología detallada, centrada en dos evaluaciones de mecánica de lectura administradas a estudiantes de Primero a Tercer grado del sistema público del país, entiéndase del Departamento de Educación de Puerto Rico (DEPR). Dichas administraciones ocurrieron en febrero y mayo de 2023.

La Fundación Flamboyán, en colaboración con el DEPR, nos facilitó dos bases de datos, una para cada administración. El total de casos en estas fuentes de información fueron los siguientes:

- Total de estudiantes en la primera administración (febrero, 2023) = 38,342
- Total de estudiantes en la segunda administración (mayo, 2023) = 35,152
- Total de escuelas participantes = 456
- Total de examinadoras(es) = 2,504

En adición, posteriormente se nos hizo llegar una base de datos adicional, la cual incluía las notas de la materia de Español de estos estudiantes a diciembre, 2023. El total de casos incluidos aquí fue de 37,468.

Para asegurar la precisión y confiabilidad de los resultados, nuestra metodología incluyó rigurosos procesos de limpieza de datos. En esta fase, eliminamos registros incompletos y duplicados, corregimos inconsistencias y consideramos casos extremos, todo lo cual aseguró una base sólida para el análisis. Además, unificamos las bases de datos en una sola y ajustamos de manera que solamente utilizáramos casos que contaran con ambas administraciones. Al finalizar la limpieza de datos, descartamos lo siguiente:

- 7,865 casos que no incluían datos de alguna de las administraciones
- 518 casos extremos¹
- 540 casos sin notas de Español

Por lo tanto, terminamos con dos cantidades totales de estudiantes para realizar nuestros análisis. En primer lugar, **al trabajar con el análisis de validación y fiabilidad de la prueba, utilizamos una muestra de 30,477 estudiantes**. Esto equivale a una reducción de 20.5% de la población original, según las bases de datos entregadas. En segundo lugar, **al trabajar con las comparativas de los resultados de la prueba de mecánica y las notas de Español, utilizamos una muestra de 29,929 estudiantes**. Esto equivale a una reducción de 22% de la población original. Considerando los grados mencionados, así como las siete regiones educativas y la distribución total de estudiantes, opinamos que esta es una muestra representativa del estudiantado de Primero a Tercero del sistema público del país.

Durante la validación de la prueba, se evaluó la consistencia interna y la correlación test-retest para verificar la fiabilidad. Además, se realizaron análisis estadísticos para examinar la validez de constructo,

¹ Entraremos en más detalles sobre los casos extremos en la siguiente sección.

comparando las puntuaciones de la prueba con indicadores externos como las notas de español, para confirmar que la prueba mide efectivamente las habilidades de lectura. Como nota técnica, dichas pruebas o análisis estadísticos fueron los siguientes: (a) prueba de Wilcoxon para conocer si las diferencias entre grupos eran estadísticamente significativas; (b) R de Cliff para conocer la magnitud o tamaño de efecto de esas diferencias; (c) correlaciones Spearman para conocer la fiabilidad temporal y la potencial relación con las calificaciones de Español. Justificamos el uso de estas pruebas al analizar varios supuestos estadísticos en comparación con las puntuaciones de todas las sub-pruebas y sub-grupos a estudiar.

ANÁLISIS DESCRIPTIVO Y RESULTADOS

En esta sección, comenzaremos explorando los datos recopilados de la prueba de mecánica de lectura. Presentaremos un desglose detallado de los resultados, ilustrando las tendencias y patrones observados en las puntuaciones de los estudiantes a lo largo de los dos periodos de evaluación. Analizamos cómo estas puntuaciones se distribuyen entre diferentes grados y regiones, proporcionando una visión clara del progreso y las áreas de mejora en las habilidades lectoras de los(las) estudiantes. Este análisis descriptivo sienta las bases para una comprensión más profunda de los impactos educativos y las implicaciones de estos hallazgos.

Variables

Entre las bases de datos que recibimos, pudimos subdividir a los(las) estudiantes por los siguientes subgrupos o variables:

- **Grados:** Primero, Segundo y Tercero.
- **Regiones educativas:** Arecibo, Bayamón, Caguas, Humacao, Mayagüez, Ponce y San Juan
- **Notas de Español** de diciembre, 2023: puntuación tradicional de A, B, C, D y F

A su vez, los datos de la prueba de mecánica de lectura se distribuían en siete (7) sub pruebas:

- Vocales y consonantes
- Sílabas
- Sílabas directas
- Sílabas inversas, trabadas, diptongos, triptongos e hiatos²
- Variantes de sonidos
- Grupos consonánticos y trabazón difícil
- Promedio de fluidez

Casos extremos

Al comenzar a observar las puntuaciones de las sub-pruebas, y durante la fase de limpieza de datos, notamos que algunos casos parecerían ser extremos en comparación con el promedio grupal. Esto significa que podrían desviarse significativamente de las demás observaciones. Por lo tanto, entendimos necesario eliminar algunos casos u observaciones. Los casos extremos pueden ser eliminados o tratados cuidadosamente por varias razones:

² En adelante, acortaremos su nombre a 'Sílabas inversas, etc.'

- **Influencia excesiva:** Pueden tener un impacto desproporcionado en los resultados estadísticos, como el promedio, distorsionando el análisis y las conclusiones.
- **Resultados no representativos:** Pueden reflejar condiciones o errores atípicos (como errores de medición o de entrada de datos) y no ser representativos de la población en estudio.
- **Modelos inadecuados:** Su presencia puede indicar que el modelo estadístico elegido no es adecuado para los datos, o que hay variables importantes no consideradas en el análisis.

Para tratar de obtener unos criterios contextualizados sobre lo que podría significar un ‘caso extremo’, nos reunimos con el personal de Fundación Flamboyán. En dicha reunión, se nos sugirió utilizar unos criterios estadísticos. Por tal razón, y como nota técnica, utilizamos el conjunto de los siguientes indicadores: (a) asimetría- simetría de la distribución de datos; (b) kurtosis- pesadez de las colas de la distribución; (c) Rango Inter-cuartil y desviaciones estándar- distribución y desviaciones de promedios; y (d) visualizaciones por medio de histogramas.

A modo de ejemplo, presentamos la distribución de puntuaciones del Primer grado (ver Figura 1) y Segundo grado (ver Figura 2) en la sub-prueba de Vocales y Consonantes de la administración ocurrida en febrero. Nótese cómo en Primer grado, la mayoría de las observaciones (frecuencias) se acumulan en las puntuaciones más bajas. Sin embargo, hay varios(as) estudiantes que obtuvieron puntuaciones sumamente altas, incluso la totalidad de las mismas. Ciertamente, se visualiza cómo estas puntuaciones extremas no son representativas de la mayoría del grupo. Por otro lado, y a nivel de grupo, las puntuaciones del Segundo grado fueron más, lo cual se aprecia con la acumulación de observaciones a mitad de la figura. Por lo tanto, y a diferencia del Primer grado, estudiantes que puntúen de manera perfecta tendrán menor influencia en ese promedio grupal. Esto es una señal de la posible necesidad de remover unos casos extremos.

Figura 1

Puntuaciones de Primer grado en sub-prueba Vocales y Consonantes de febrero

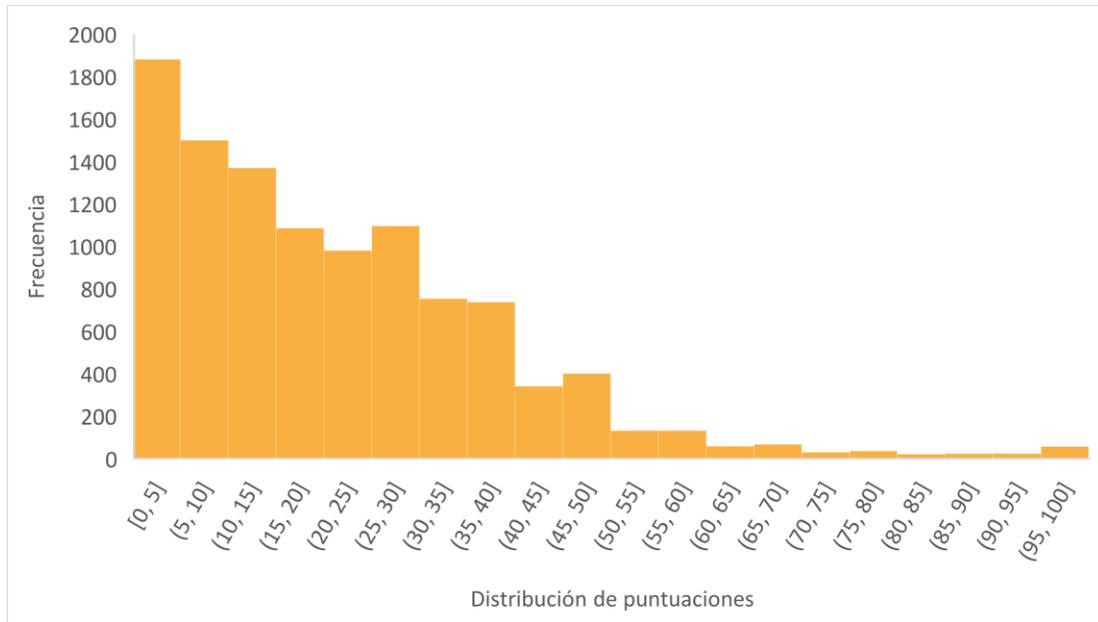
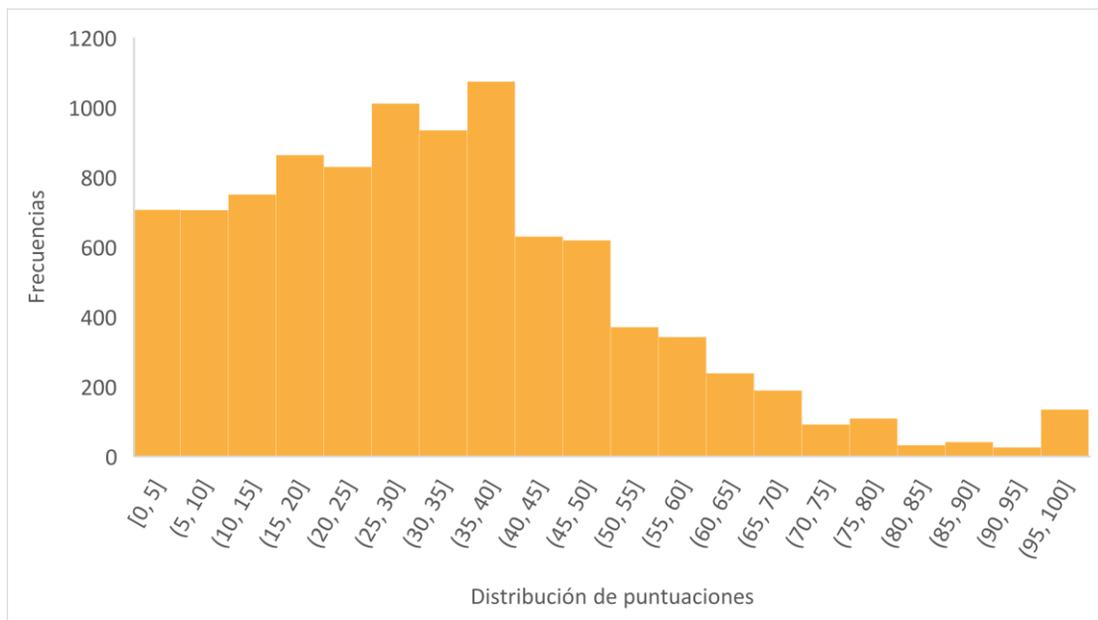


Figura 2

Puntuaciones de Segundo grado en sub-prueba Vocales y Consonantes de febrero



Ahora bien, ¿qué casos extremos deben ser removidos de manera que no acabemos con influencias excesivas y promedios no representativos del grupo mayor? Es importante notar que no hay un punto de corte pre-determinado, sino que este debe ocurrir de manera informada con los indicadores que mencionamos previamente, así como con la contextualización de la Educación como disciplina y la experiencia en enseñanza de la lectura en grados primarios.

Al no obtener el detalle sobre la contextualización en Educación y la enseñanza de la lectura, realizamos unos cortes con los indicadores estadísticos. Los mismos fueron los siguientes:

- **Primer corte:** en Primer grado, prueba Vocales y Consonantes de febrero, se eliminaron aquellas observaciones cuyas puntuaciones fueron de 80 o más (n = 148)
- **Segundo corte:** Primer grado, prueba Sílabas de febrero, se eliminan aquellas observaciones cuyas puntuaciones fueron de 80 puntos o más (n = 150)
- **Tercer corte:** Primer grado, prueba Promedio de Fluidez, se eliminan aquellas observaciones cuyas puntuaciones fueron de 75 puntos o más (n = 101)
- **Cuarto corte:** Segundo grado, prueba Promedio de Fluidez, se eliminan aquellas observaciones cuyas puntuaciones fueron de 130 puntos o más (n = 119)

Para hacer justicia tanto al grupo en su totalidad, así como a posibles casos extremos cuyas puntuaciones fueron reales y bien merecidas, decidimos realizar unos cortes sumamente conservadores. Esto quiere decir que es posible extenderlos, si así lo determina la experiencia en la enseñanza de la lectura en grados primarios.

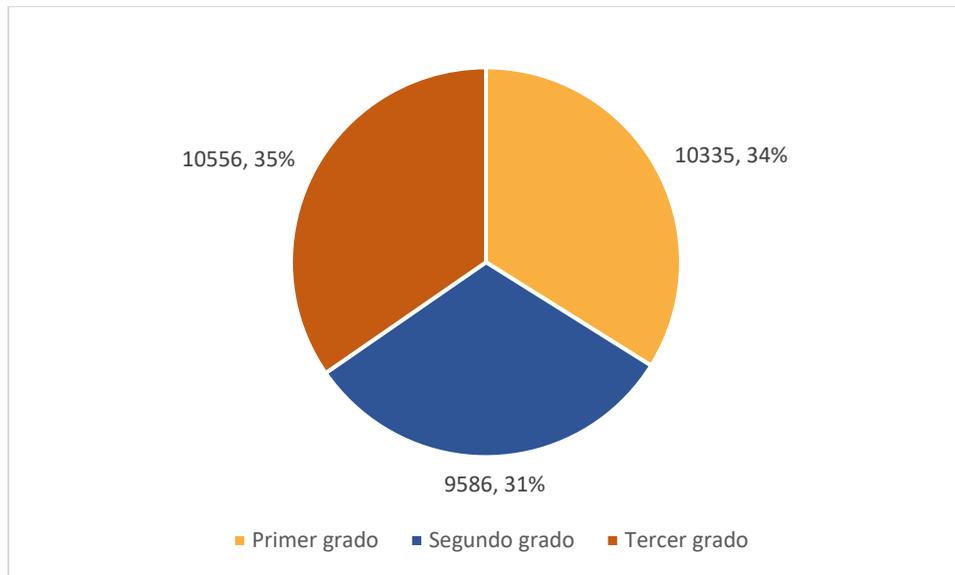
Del análisis anterior se desprende que **la mayoría de los casos extremos provienen del Primer grado**. Encontramos una gran variabilidad en las puntuaciones de las sub-pruebas para este grado, así como en sus calificaciones de Español. ¿Por qué esto es así? ¿Es un asunto que viene del kínder o la educación pre-escolar? Además, encontramos menor variabilidad en los grados posteriores; entonces, ¿significa que se corrige en la educación escolar? En definitiva, aquí hay materia para futuras investigaciones.

Distribución de estudiantes

Luego de la limpieza de datos, observamos las distribuciones de estudiantes según los sub-grupos. Al subdividirles por grado (ver Figura 3), vemos que el Primer grado cuenta con representación de 10,335 estudiantes, o el 34% de la muestra; el Segundo grado contaba con 9,586 (31%); y el Tercer grado con 10,556 (35%). Luego de un análisis estadístico, **no encontramos desproporciones significativas por grado**.

Figura 3

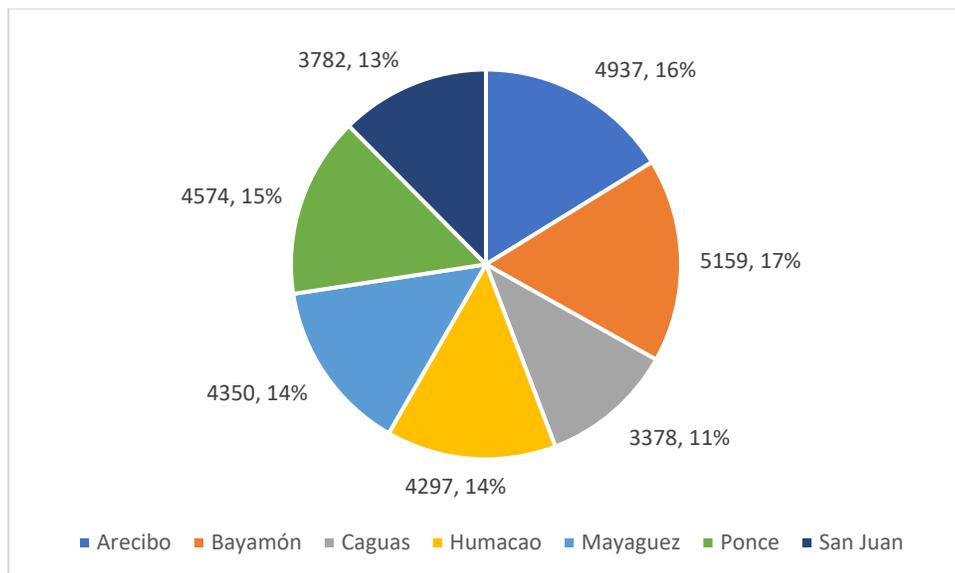
Distribución estudiantil por grado



A su vez, **la distribución por región educativa tampoco tuvo desproporciones significativas** (ver Figura 4). Se pueden observar las siguientes frecuencias y porcentajes: Arecibo con 4,937 (16%), Bayamón con 5,159 (17%), Caguas con 3,378 (11%), Humacao con 4,297 (14%), Mayagüez 4,350 (14%), Ponce con 4,574 (15%) y San Juan con 3,782 (13%).

Figura 4

Distribución estudiantil por región educativa



Descriptivos³

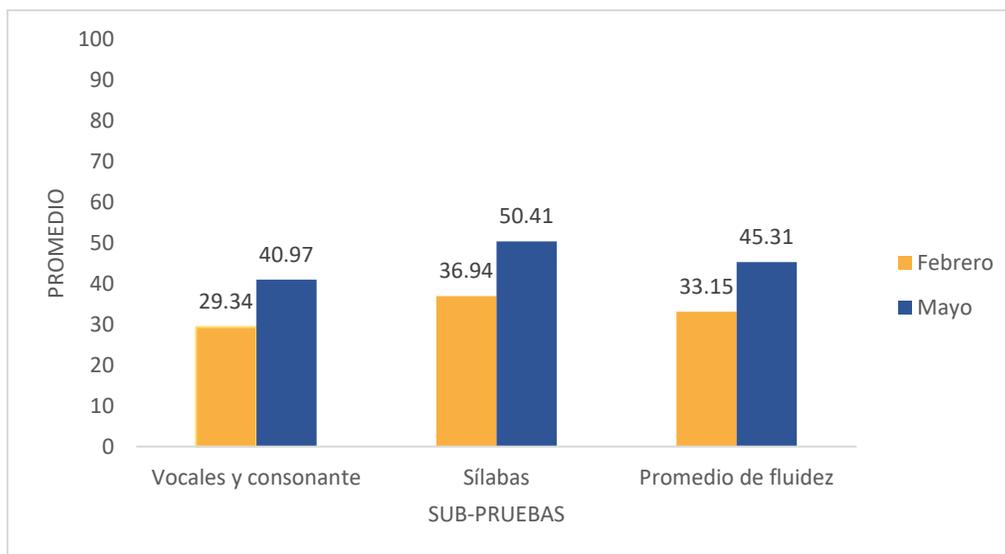
Al conocer los límites de nuestra población, pasamos a realizar diferentes análisis descriptivos como medidas de tendencia central, medidas de dispersión, distribución de datos, así como sus visualizaciones. El motivo es poder comparar las puntuaciones de ambas administraciones de la prueba, para así dar cuenta de su potencial utilidad como instrumento que mide el desempeño estudiantil en aquello que pretende medir- entiéndase, la mecánica de lectura. A continuación, presentamos estos resultados.

Descriptivos a nivel grupal

En primer lugar, y **a nivel grupal, pudimos observar un aumento de promedios para todas las sub-pruebas** (ver Figura 5 y Figura 6)⁴. El hecho de que los promedios hayan aumentado entre febrero y mayo sugiere una mejora general en el rendimiento de los estudiantes. En otras palabras, el cambio en promedio sugiere que los(las) estudiantes respondieron de manera positiva a las intervenciones educativas, que se generó aprendizaje y que el mismo influyó en las destrezas de mecánica de lectura.

Figura 5

Promedios de sub-pruebas Vocales y Consonantes, Sílabas y Promedio de Fluidez para todos los grupos y en ambas administraciones

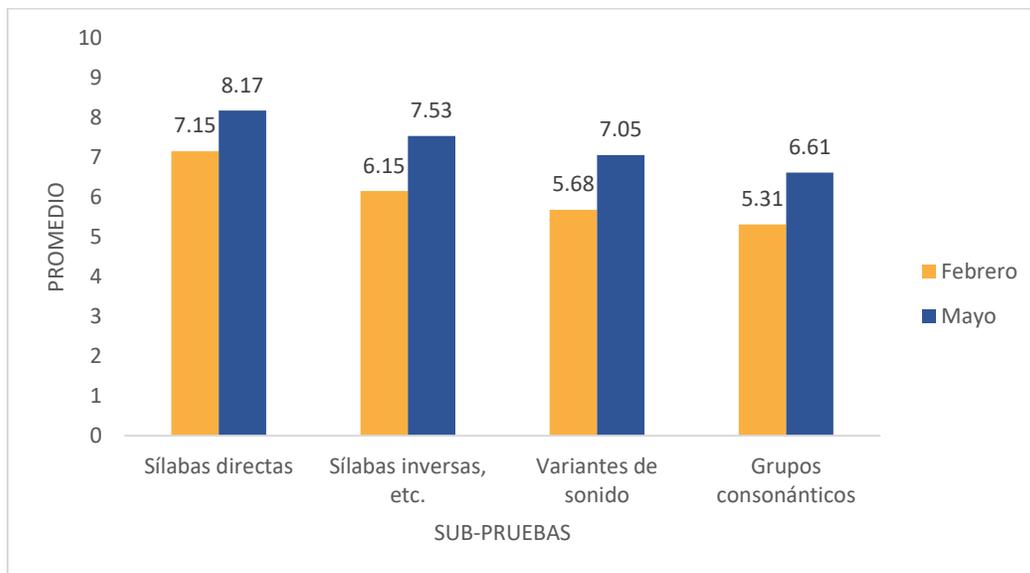


³ Ver Anejo para datos adicionales.

⁴ Como nota de edición, resaltamos que unimos en una sola figura aquellas pruebas que tenían una puntuación máxima de 100 (Vocales y Consonantes; y Sílabas) o más (Promedio de Fluidez), y las separamos de aquellas sub-pruebas cuya puntuación máxima era de 10. Esto lo hicimos para no generar una percepción errónea de que los(las) estudiantes salieron mejor o peor en unas sub-pruebas versus otras.

Figura 6

Promedios de sub-pruebas Sílabas directas, Sílabas inversas, etc., Variantes de sonido y Grupos consonánticos para todos los grupos y en ambas administraciones



Estos promedios también sirven de línea base para futuras administraciones de la prueba. Aunque sería ideal administrar la misma en tiempos aún más alejados (por ejemplo, y como mínimo, comenzando en septiembre y culminando en mayo), estos resultados pueden ser utilizados para comparativas longitudinales.

Aunque la prueba posee un contenido distinto a nivel de grados, estas figuras nos permiten visualizar el desarrollo o desempeño estudiantil a nivel de grados primarios como conjunto. Con las figuras también podemos ir visualizando la consistencia de la prueba al tratar de medir en dos tiempos distintos lo que llamamos mecánica de lectura. Sin embargo, **a la hora de tomar decisiones individualizadas y de estrategias educativas, sugerimos utilizar el análisis a nivel de grados**, pues hace más justicia a la a sus contenidos únicos.

Descriptivos a nivel de grados

A su vez, **a nivel de grados también encontramos aumento de promedios para todas las sub-pruebas** (ver Figuras 7 a la 12). Por lo tanto, las implicaciones fueron las mismas que aquellas que observamos a nivel grupal: los(las) estudiantes respondieron de manera positiva a las intervenciones educativas, que se generó aprendizaje y que el mismo influyó en las destrezas de mecánica de lectura. Además, estos promedios también sirven de línea base para futuras administraciones de la prueba.

Figura 7

Promedios de sub-pruebas Vocales y Consonantes, Sílabas y Promedio de Fluidez para el Primer grado y en ambas administraciones

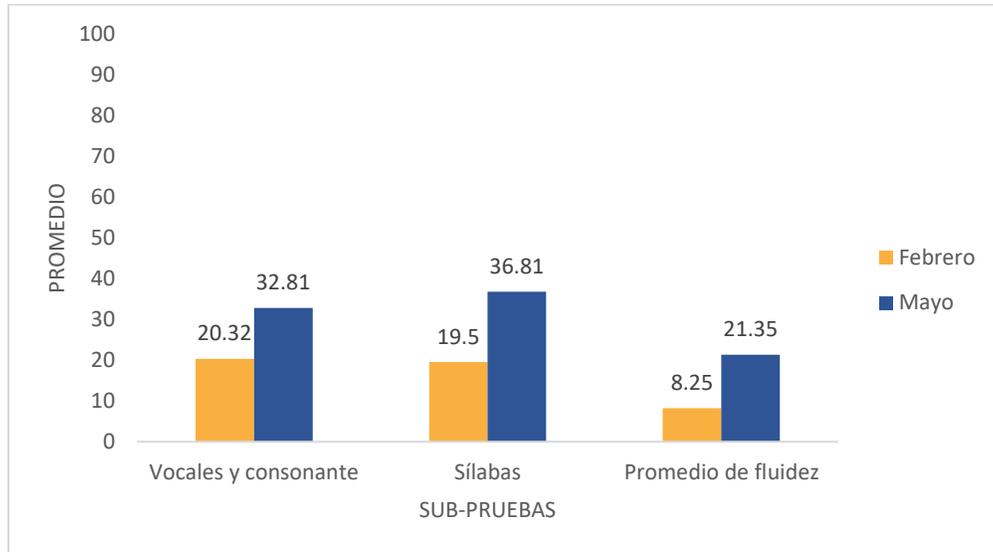


Figura 8

Promedios de sub-pruebas Sílabas directas, Sílabas inversas, etc., Variantes de sonido y Grupos consonánticos para el Primer grado y en ambas administraciones

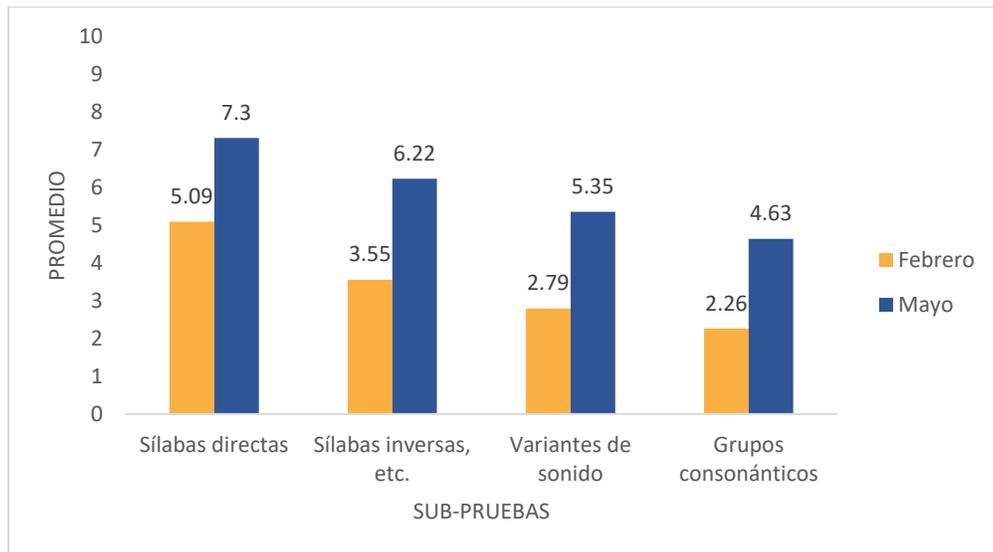


Figura 9

Promedios de sub-pruebas Vocales y Consonantes, Sílabas y Promedio de Fluidez para el Segundo grado y en ambas administraciones

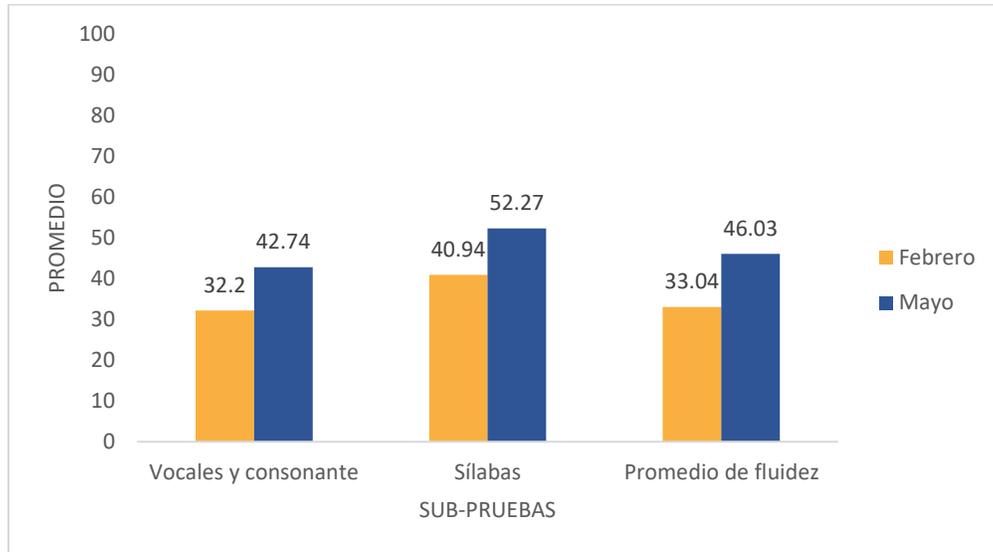


Figura 10

Promedios de sub-pruebas Sílabas directas, Sílabas inversas, etc., Variantes de sonido y Grupos consonánticos para el Segundo grado y en ambas administraciones

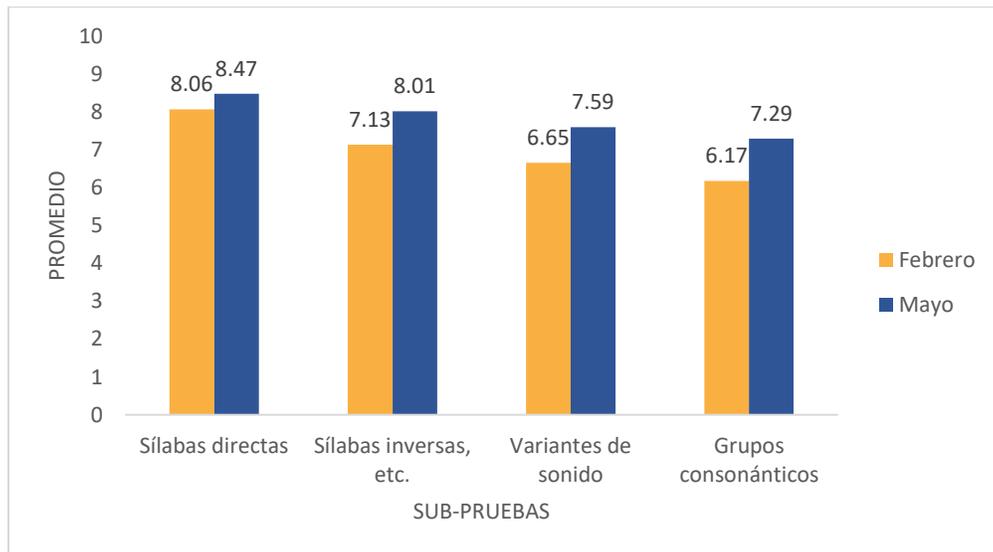


Figura 11

Promedios de sub-pruebas Vocales y Consonantes, Sílabas y Promedio de Fluidez para el Tercer grado y en ambas administraciones

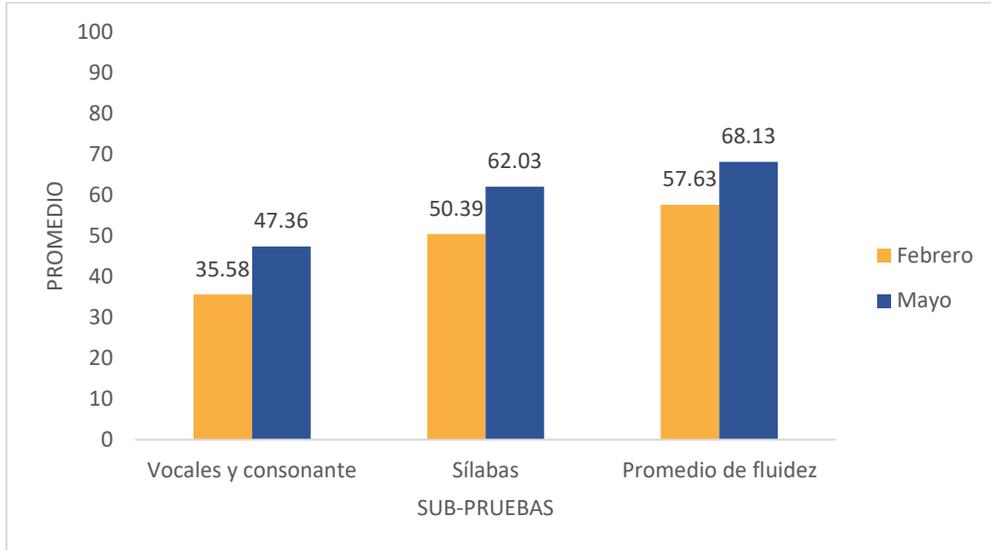
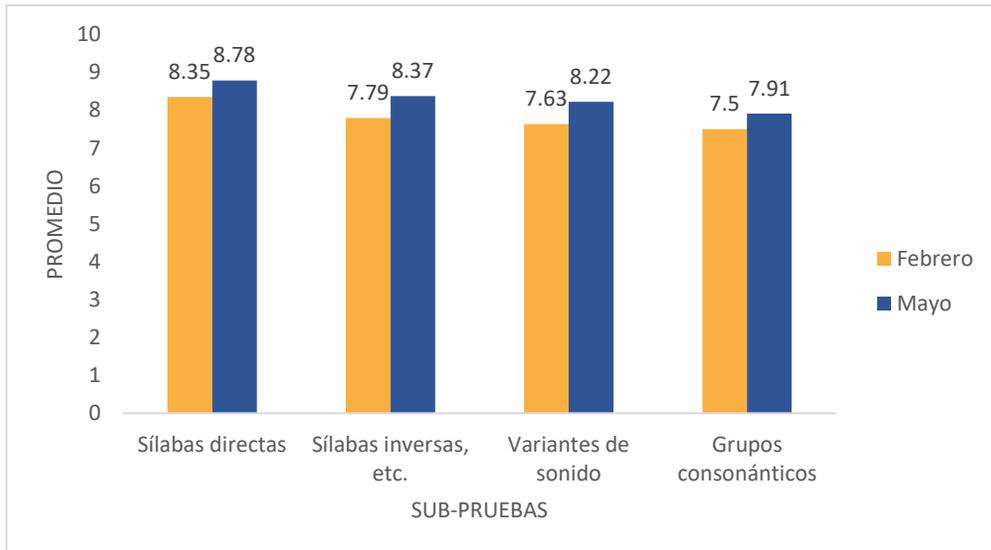


Figura 12

Promedios de sub-pruebas Sílabas directas, Sílabas inversas, etc., Variantes de sonido y Grupos consonánticos para el Tercer grado y en ambas administraciones



Otro hallazgo importante es que no solamente ocurrió el aumento entre ambos tiempos a nivel de grado, sino que **también hubo un incremento de promedios proporcional al grado**. Entiéndase, el

Segundo grado obtuvo promedios superiores a los de Primer grado; y, a su vez, el Tercer grado obtuvo promedios superiores a los del Segundo grado. Por lo tanto, podemos notar cómo se van dominando estas destrezas en la mecánica de lectura a medida que la población pasa a su próximo nivel.

También cabe señalar que, en proporción, los(las) estudiantes parecen obtener mejores puntuaciones para las sub-pruebas de Sílabas directas, Sílabas inversas, etc., Variantes de sonido y Grupos consonánticos, que para las sub-pruebas de Vocales y consonantes, Sílabas y Promedio de Fluidez. Por lo tanto, se podría sugerir decir que dominan mejor ese contenido. Mencionamos esto, pues las primeras que mencionamos se evidencia un promedio más cercano al máximo de puntuación. Sin embargo, cabe resaltar que cada sub-prueba ocurre en un minuto. El segundo conjunto de sub-pruebas tiene mucho más contenido que el primer conjunto, por lo que esa conclusión, en nuestra opinión, no se sostiene. Por ejemplo, en base a esa lógica, sería erróneo sugerir que un(a) estudiante domina más una prueba de Sílabas directas que una de Vocales y consonantes.

Además, y como señalamos anteriormente, a la hora de tomar decisiones individualizadas y de estrategias educativas, sugerimos utilizar el análisis a nivel de grados, pues hace más justicia a la a sus contenidos únicos. Sin embargo, **estos resultados también deben ser interpretados en el contexto del programa educativo y las características específicas de los(las) estudiantes**. Otros factores como el contenido de las lecciones, las metodologías de enseñanza y el entorno del aula pueden influir en el aprendizaje.

Por último, también quedan preguntas por contestar y que son motivo de futuras investigaciones. Algunos ejemplos son las siguientes. ¿Un(a) experto(a) en la enseñanza de la lectura en grados primarios esperaría promedios similares para los meses en que administraron la prueba? ¿Observaría fácilmente un rezago académico o, al contrario, buenas señales de rendimiento académico? ¿Se justifica el aumento gradual de promedios a nivel de grados? ¿Podemos concluir que los resultados se encuentran a nivel de grado? En el caso de interpretar los resultados como rezago, ¿qué otros factores podrían influir en estas puntuaciones de manera que podamos impactar en ellas? En el caso de que no interpretemos los resultados como rezago, ¿cuáles son las fortalezas de los(las) estudiantes, las estrategias de enseñanza que mejor funcionan y componentes comunitarios que apoyan al desempeño? Al abordar estas preguntas, se profundiza en la comprensión del aprendizaje en lectura, permitiendo desarrollar estrategias educativas más efectivas y contribuir significativamente al desarrollo académico y al progreso nacional en la educación.

Descriptivos a nivel regional

Al subdividir los promedios por regiones educativas, también encontramos aumento de promedios para todas las sub-pruebas (ver Tablas 1 y 2). Lo mismo ocurrió al subdividir por cada grado, según región educativa⁵. Por lo tanto, las implicaciones fueron las mismas que aquellas que observamos a nivel grupal y al subdividir por grados. Además, podemos destacar que, mientras hay diferencias en los promedios entre las distintas regiones, las mismas deben interpretarse con cautela. Aunque utilizamos pruebas estadísticas para estudiar las diferencias, debido al gran tamaño de la muestra, estas pruebas pueden detectar diferencias pequeñas como estadísticamente significativas. Por lo tanto, futuras

⁵ Ver Anejo para datos adicionales.

investigaciones pudieran indagar a mayor profundidad- utilizando pruebas de significancia estadística, tamaños de efecto, entre otras- para evaluar la relevancia práctica de las diferencias observadas.

Tabla 1

Promedios de sub-pruebas Vocales y Consonantes, Sílabas y Promedio de Fluidez para todos los grados y en ambas administraciones

Regiones educativas	Vocales y Consonantes (febrero)	Vocales y Consonantes (mayo)	Sílabas (febrero)	Sílabas (mayo)	Promedio de fluidez (febrero)	Promedio de fluidez (mayo)	N
Arecibo	27.90	40.31	35.34	49.25	31.91	43.51	4937
Bayamón	28.51	39.28	38.24	52.62	34.99	46.82	5159
Caguas	32.58	44.83	38.63	52.65	34.89	47.93	3378
Humacao	29.59	42.65	36.25	50.25	31.32	44.21	4297
Mayagüez	29.98	41.37	35.42	48.28	31.68	43.81	4350
Ponce	27.75	38.19	36.73	49.11	33.23	45.55	4574
San Juan	30.38	41.70	38.55	51.09	34.38	45.97	3782

Tabla 2

Promedios de sub-pruebas Sílabas directas, Sílabas inversas, etc., Variantes de sonido y Grupos consonánticos para todos los grados y en ambas administraciones

Regiones Educativas	Sílabas directas (febrero)	Sílabas directas (mayo)	Sílabas inversas (febrero)	Sílabas inversas (mayo)	Variantes de sonido (febrero)	Variantes de sonido (mayo)	Grupos consonánticos (febrero)	Grupos consonánticos (mayo)
Arecibo	7.07	8.20	6.04	7.53	5.53	7.04	5.23	6.59
Bayamón	7.40	8.53	6.40	7.90	5.89	7.36	5.57	6.93
Caguas	7.46	8.24	6.52	7.61	6.01	7.21	5.66	6.73
Humacao	7.11	8.26	5.96	7.57	5.60	7.07	5.14	6.57
Mayagüez	6.88	7.82	5.93	7.18	5.52	6.73	5.09	6.30
Ponce	7.09	8.19	6.11	7.49	5.63	6.98	5.25	6.55
San Juan	7.07	7.94	6.10	7.33	5.64	6.92	5.22	6.53

Nota. Los tamaños de la muestra son idénticos a los presentados en la Tabla 1.

En resumen, el análisis descriptivo reveló un incremento significativo en los promedios de las pruebas de mecánica de lectura de febrero a mayo. Esta mejora se observó tanto en el análisis general como al desglosar los datos por grados y regiones educativas. Las diferencias en las puntuaciones se confirmaron mediante pruebas estadísticas, mostrando la utilidad de la prueba para medir el desarrollo de las habilidades de mecánica de lectura.

Análisis de validez y fiabilidad

En esta sección abordaremos aspectos cruciales para entender la efectividad de la prueba de mecánica de lectura. Iniciaremos explorando las diferencias significativas entre los resultados de febrero y mayo, buscando evidencia de mejoras en el aprendizaje. Posteriormente, nos enfocaremos en el tamaño del efecto de estas diferencias para evaluar su importancia práctica. Finalmente, analizaremos las correlaciones entre ambos tiempos, lo cual nos permitirá evaluar la consistencia y fiabilidad de la prueba a lo largo del tiempo.

Análisis de significancia estadística

Para evaluar las diferencias significativas entre los resultados de febrero y mayo de la prueba de mecánica de lectura, se utilizó la prueba de Wilcoxon. Esta prueba es adecuada para muestras con propiedades no paramétricas, como sucedió con la población estudiantil. Este enfoque no paramétrico nos permite una evaluación más precisa de si las mejoras observadas son estadísticamente significativas, independientemente de la distribución normal de los datos.

Todas las pruebas mostraron un valor de $p < .001$, lo que indica que las diferencias en el rendimiento entre los dos tiempos de la prueba son estadísticamente significativas. Esto sugiere que las mejoras observadas en las habilidades de lectura de los estudiantes no son producto del azar, sino el resultado de procesos educativos o cambios significativos en su aprendizaje durante el periodo evaluado.

La significancia estadística encontrada refuerza la utilidad de la prueba de mecánica de lectura como una herramienta confiable para medir habilidades lectoras. Indica que los cambios observados en las puntuaciones son sustanciales y consistentes, lo que sugiere que **la prueba es efectiva en capturar las variaciones genuinas en la mecánica de lectura de los(las) estudiantes a lo largo del tiempo.** Esto valida su aplicabilidad en entornos educativos para evaluar y monitorear el progreso en la lectura.

Tamaño de efecto

En el análisis subsiguiente, nos centramos en los tamaños del efecto, utilizando la prueba r de Cliff, para cuantificar la magnitud de las diferencias observadas entre los dos tiempos de la prueba de mecánica de lectura. La r de Cliff se utiliza porque es adecuada para medir el tamaño del efecto en datos no paramétricos (Cliff, 1993), como los obtenidos para los(las) estudiantes. Es útil cuando no podemos asumir una distribución normal de los datos y cuando las diferencias en las medianas son más relevantes que las diferencias en las medias. La r de Cliff proporciona una medida clara de cuánto se superponen las distribuciones de dos grupos, lo que es crucial para entender el impacto real de las intervenciones educativas en las habilidades de lectura, ya asumiendo la validez de la prueba de su mecánica.

Los resultados de la *r* de Cliff para la prueba de mecánica de lectura indican variaciones en la magnitud de las diferencias entre los tiempos de prueba, por grado y sub-prueba (ver Tabla 3). En primer grado, se observan efectos medianos en la mayoría de las sub-pruebas, sugiriendo cambios más notables en habilidades lectoras. En segundo y tercer grado, los efectos varían desde insignificantes a pequeños, indicando diferencias menos pronunciadas. Estos resultados reflejan la evolución y los desafíos del aprendizaje lector a lo largo de los grados, resaltando áreas específicas donde los(las) estudiantes muestran distintos niveles de progreso.

Tabla 3

Resultados de R de Cliff (Delta) por Grado

Grados	Sub-pruebas	Delta	Límite inferior	Límite superior	Magnitud
Primer grado	Vocales y consonantes	-.35	-.37	-.34	Mediana
	Sílabas	-.39	-.40	-.37	Mediana
	Sílabas directas	-.32	-.34	-.31	Pequeña
	Sílabas inversas	-.36	-.37	-.34	Mediana
	Variantes de sonido	-.36	-.37	-.34	Mediana
	Grupos consonánticos	-.33	-.34	-.31	Pequeña
	Promedio de Fluidez	-.36	-.38	-.35	Mediana
Segundo grado	Vocales y consonantes	-.27	-.29	-.26	Pequeña
	Sílabas	-.23	-.25	-.22	Pequeña
	Sílabas directas	-.08	-.10	-.07	Insignificante
	Sílabas inversas	-.12	-.14	-.11	Insignificante
	Variantes de sonido	-.21	-.22	-.19	Pequeña
	Grupos consonánticos	-.21	-.22	-.19	Pequeña
	Promedio de Fluidez	-.23	-.25	-.21	Pequeña
Tercer grado	Vocales y consonantes	-.29	-.30	-.27	Pequeña
	Sílabas	-.25	-.27	-.24	Pequeña
	Sílabas directas	-.12	-.14	-.11	Insignificante
	Sílabas inversas	-.13	-.14	-.11	Insignificante
	Variantes de sonido	-.14	-.15	-.12	Insignificante
	Grupos consonánticos	-.05	-.06	-.04	Insignificante
	Promedio de Fluidez	-.17	-.18	-.15	Pequeña

Nota. La magnitud se basa en un criterio clásico de Cohen (1988). Aunque el criterio sea útil y pueda ser entendido fácilmente, debe ser contextualizado para conocer bien sus implicaciones prácticas.

Como nota técnica, y para ayudar en la interpretación, el signo negativo (-) indica que la tendencia en los datos es tal que el primer grupo (promedios de febrero) tiende a tener valores más bajos en comparación con el segundo grupo (promedios de mayo). En otras palabras, **hay una tendencia hacia valores más altos en la segunda medición**. Además, los valores de límite inferior y superior implican un intervalo de confianza. Esto indica que, con un 95% de confianza, el verdadero valor de la *r* de Cliff en la

población general está dentro de este rango. El hecho de que este intervalo no incluya el 0 refuerza la idea de que la diferencia observada no es producto del azar.

Los resultados obtenidos de la r de Cliff para la prueba de mecánica de lectura concuerdan con las expectativas educativas. Es razonable anticipar que los(las) estudiantes en primer grado, cuyas lecciones se centran intensamente en la mecánica de lectura, muestren una mayor similitud entre sus habilidades y el contenido evaluado en la prueba. Por el contrario, en grados superiores, como segundo y tercero, donde el currículo se expande y se vuelve más complejo, las puntuaciones tienden a divergir más del contenido específico de la prueba. **Esta disminución gradual en los tamaños del efecto no solo refleja un proceso de aprendizaje natural y progresivo en la mecánica de lectura, sino que también valida la efectividad de la prueba en medir lo que se propone.** Estos hallazgos subrayan la relevancia de la prueba como un indicador fiable del nivel de dominio de los estudiantes en habilidades lectoras fundamentales, adaptándose adecuadamente a su desarrollo académico a lo largo de los años.

Por último, **es importante considerar estos resultados en el contexto del estudio y el campo de investigación para confirmar la relevancia práctica de esta diferencia.** Por ejemplo, en algunas áreas, una diferencia de esta magnitud podría ser muy significativa, mientras que en otras podría no serlo. ¿Cómo ocurre en el campo de la Educación para pruebas similares? A tales propósitos, realizamos una revisión de la literatura en el campo de la educación, en donde también hayan utilizado la r de Cliff. En ella hemos constatado que los tamaños del efecto obtenidos en nuestro estudio son comparables a los reportados en investigaciones del campo (DeWilde, Brysbaert y Eyckmans, 2022; McPadden y Brewes, 2017; Nielsen, 2023; Palmgren, Tuominen y Kontro, 2022). Este hallazgo provee un fundamento para corroborar la relevancia y significancia de nuestros resultados, indicando que las diferencias observadas en las habilidades de lectura de los estudiantes no solo son estadísticamente significativas, sino también significativas desde un punto de vista práctico y educativo. La similitud de nuestros tamaños del efecto con los de estudios previos en educación refuerza la validez de nuestras conclusiones y su aplicabilidad en el contexto educativo más amplio, ofreciendo una perspectiva valiosa para futuras intervenciones y estrategias pedagógicas.

Fiabilidad temporal

Para evaluar la fiabilidad temporal de la prueba de mecánica de lectura, llevamos a cabo correlaciones entre las puntuaciones obtenidas por los(las) estudiantes en los dos momentos de evaluación, febrero y mayo. Este análisis es fundamental para confirmar la confiabilidad de la prueba, ya que una alta correlación test-retest indica que la prueba mide consistentemente las habilidades de lectura a lo largo del tiempo. Además, esta consistencia es un componente crucial para validar la prueba, demostrando que no solo es confiable, sino que también mide de forma efectiva y repetible el constructo deseado. Al asegurarnos de que nuestra herramienta de evaluación es tanto válida como confiable, podemos confiar en que los resultados obtenidos proporcionan una representación precisa del desarrollo de habilidades de lectura en los estudiantes.

En nuestro estudio, optamos por la prueba de correlaciones de Spearman debido a las características específicas de nuestra población y los datos recopilados. Dado que los datos no seguían una distribución normal, la prueba de Spearman, que no asume normalidad, fue la más adecuada para evaluar la relación entre las variables de interés.

Al evaluar la prueba de mecánica de lectura, las correlaciones entre las puntuaciones de los dos tiempos de prueba revelaron consistencias notables en todas las sub-pruebas y grados (ver Tabla 4). En primer grado, las correlaciones oscilaron entre 0.65 y 0.81, destacando una fuerte consistencia en habilidades como 'Sílabas'. En segundo grado, las correlaciones se mantuvieron altas, variando de 0.65 a 0.83, con una consistencia particularmente destacada en 'Promedio de Fluidez'. En tercer grado, aunque las correlaciones fueron ligeramente más bajas, de 0.55 a 0.78, aún indicaron una buena consistencia en las habilidades de lectura, especialmente en 'Promedio de Fluidez' y 'Sílabas'.

Tabla 4

Correlaciones entre ambas administraciones de la prueba de mecánica de lectura, según grado

Grados	Sub-pruebas	Correlación entre tiempos
Primer grado	Vocales y consonantes	.72
	Sílabas	.81
	Sílabas directas	.71
	Sílabas inversas	.71
	Variantes de sonido	.70
	Grupos consonánticos	.65
	Promedio de Fluidez	.69
Segundo grado	Vocales y consonantes	.71
	Sílabas	.81
	Sílabas directas	.65
	Sílabas inversas	.68
	Variantes de sonido	.68
	Grupos consonánticos	.70
	Promedio de Fluidez	.83
Tercer grado	Vocales y consonantes	.67
	Sílabas	.77
	Sílabas directas	.55
	Sílabas inversas	.59
	Variantes de sonido	.57
	Grupos consonánticos	.63
	Promedio de Fluidez	.78

Nota. Todas las correlaciones obtuvieron un $p < .001$ y son clasificadas como altas, según los estándares tradicionales (Cohen, 1988).

A modo de explicación, si tenemos un número cercano a 1, significa que hay una fuerte relación entre ambas administraciones; por ejemplo, cuando una sube, la otra también tiende a subir, y viceversa. Si el número es cercano a 0, significa que no hay mucha relación. En nuestro caso, los números de correlación están entre 0.55 y 0.83, lo que significa que hay una relación bastante fuerte entre cómo los(las) estudiantes puntuaron en la primera y la segunda prueba. Cuanto más alto es el número, más fuerte es

esa relación. Por ejemplo, un 0.81 indica que los estudiantes que lo hicieron bien en la primera prueba, generalmente también lo hicieron bien en la segunda.

Estos resultados sugieren que la prueba de mecánica de lectura es un instrumento fiable y consistente a lo largo del tiempo, reflejando eficazmente el desarrollo de habilidades lectoras en los(las) estudiantes. La fuerte correlación test-retest en las sub-pruebas subraya su validez para medir las habilidades de mecánica de lectura, ofreciendo una herramienta valiosa para evaluar y monitorear el progreso lector en los primeros años de la educación primaria.

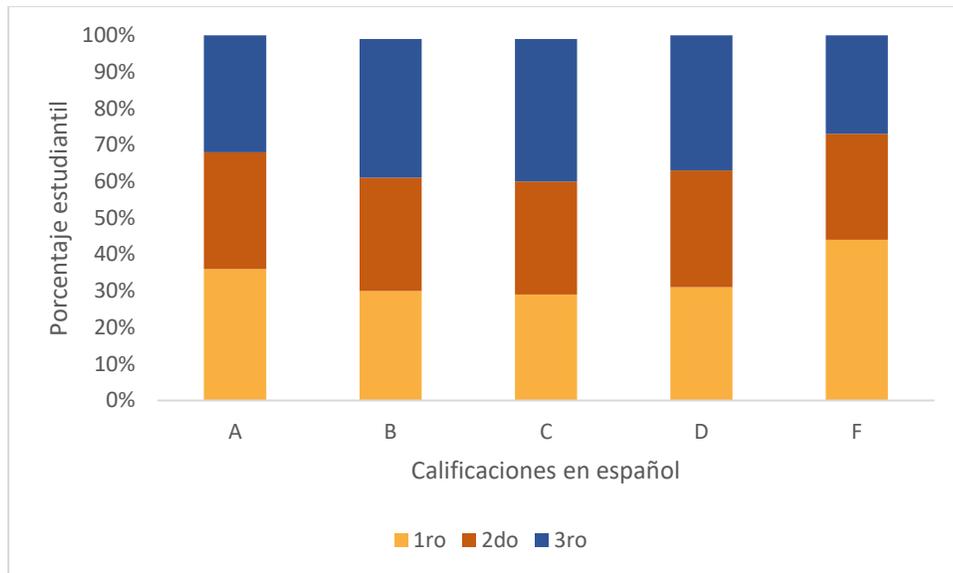
Comparación de la prueba de mecánica de lectura con notas de Español

En la próxima sección del análisis, exploraremos la relación entre los promedios de las pruebas de mecánica de lectura y las calificaciones de español de los(las) estudiantes a diciembre de 2023. Esta comparación es crucial para estudiar la validez y confiabilidad de nuestra prueba de mecánica de lectura, ya que nos permite entender si las habilidades que mide la prueba están efectivamente relacionadas con el rendimiento académico general de los(las) estudiantes en español, materia en que principalmente se enseña sobre la mecánica de la lectura. Al correlacionar estas dos métricas, podemos verificar si la prueba es un indicador fiable del éxito académico en áreas relacionadas con la lectura, lo cual es esencial para confirmar su utilidad y precisión como herramienta de evaluación educativa.

Comenzamos primero mostrando un desglose de las calificaciones en español para tres grados distintos: primero, segundo y tercero (ver Figura 13). Cada barra representa el porcentaje de estudiantes que obtuvieron una determinada calificación (A, B, C, D, F), con los colores diferenciando los grados. Al observar las barras, notamos que **una mayor proporción de estudiantes de tercer grado obtiene calificaciones A y B en comparación con los de primero y segundo.** Por otro lado, **las calificaciones D y F son más comunes en primer grado** que en los otros dos. Esto podría sugerir varios asuntos: que a medida que los(las) estudiantes avanzan en su educación, tienden a obtener mejores calificaciones en español; que los criterios de evaluación varían por grado; que se obtiene mucha mayor variabilidad en la experiencia educativa previa al primer grado, en comparación con segundo y tercero; entre otras.

Figura 13

Distribución de notas de español (diciembre, 2023), según grados



A continuación, presentamos unas figuras que reflejan la relación entre las calificaciones de Español y los promedios en distintas subpruebas de la evaluación de mecánica de lectura (ver Figuras 14 a la 17). En general, se observa una tendencia descendente: **los(las) estudiantes con calificaciones más altas en español tienden a tener promedios más altos en las subpruebas de mecánica de lectura**. Esto era de esperar, pues partíamos del siguiente supuesto: si la prueba de lectura tiene contenidos similares o idénticos a aquellos de la clase de Español, entonces, su relación debía ser lineal y descendiente-entiéndase, a menor la nota de Español, menor el promedio en la prueba de mecánica de lectura. Esto puede indicar que las **habilidades de lectura evaluadas en las subpruebas están relacionadas con el desempeño académico en Español, y sugiere que la prueba de mecánica de lectura podría ser un reflejo fiable del conocimiento y las habilidades lingüísticas que los(las) estudiantes están aprendiendo en sus clases de Español**. La similitud en las tendencias entre febrero y mayo podría sugerir que estas habilidades se mantienen o incluso se fortalecen con el tiempo, reflejando la posible influencia de la instrucción educativa durante el semestre. Dicho análisis y sus conclusiones también se sostienen al subdividir los resultados por grados⁶.

⁶ Ver Anejo para datos adicionales.

Figura 14

Relación entre sub-pruebas de febrero- Vocales y Consonantes, Sílabas y Promedio de Fluidez- y notas de español (diciembre, 2023) para todos los grupos

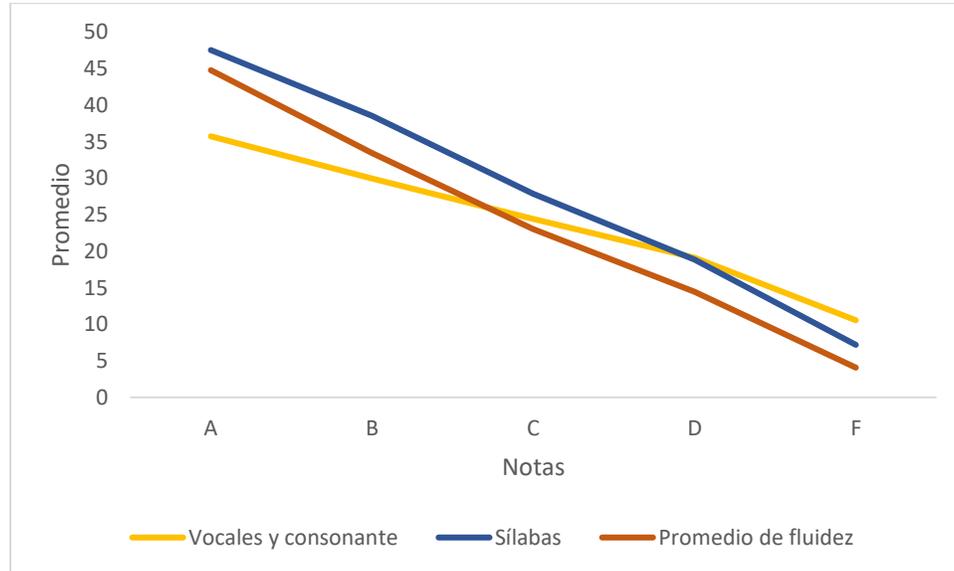


Figura 15

Relación entre sub-pruebas de mayo- Vocales y Consonantes, Sílabas y Promedio de Fluidez- y notas de español (diciembre, 2023) para todos los grupos

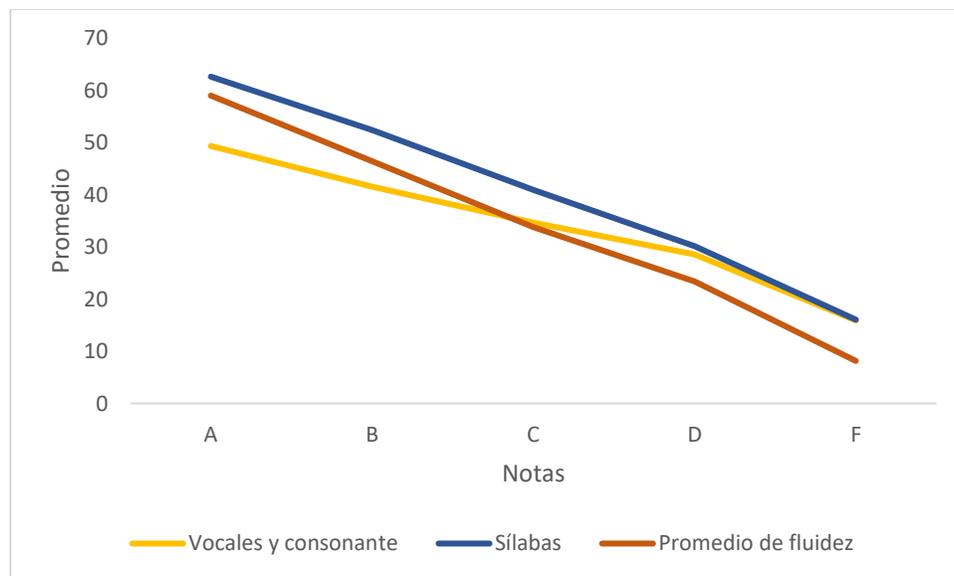


Figura 16

Relación entre sub-pruebas de febrero- Sílabas directas, Sílabas inversas, etc., Variantes de sonido y Grupos consonánticos- y notas de español (diciembre, 2023) para todos los grupos

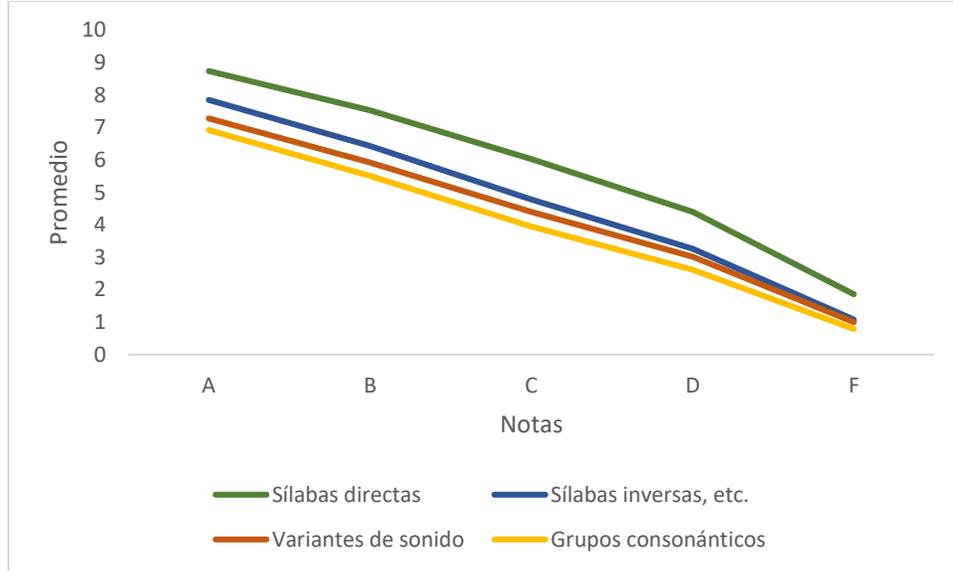
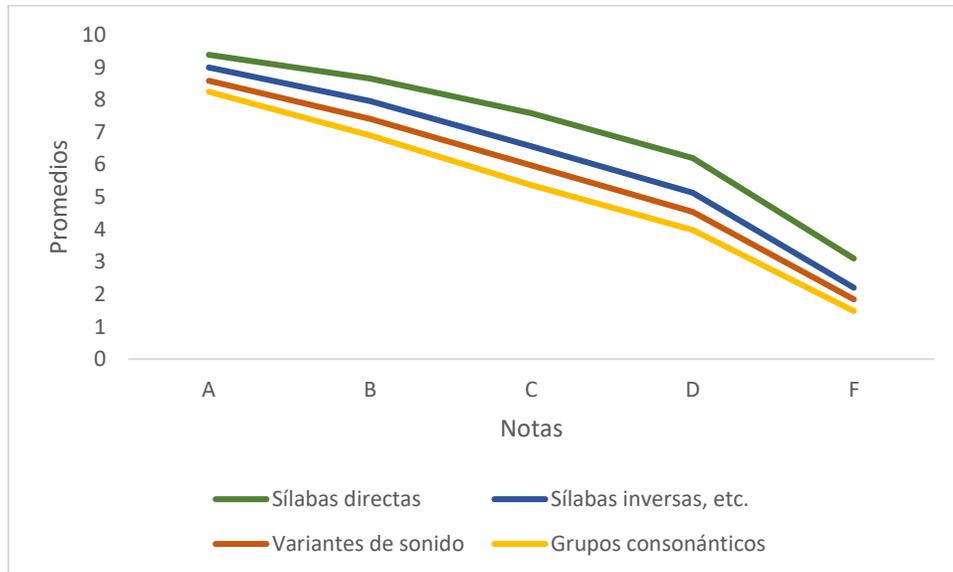


Figura 17

Relación entre sub-pruebas de mayo- Sílabas directas, Sílabas inversas, etc., Variantes de sonido y Grupos consonánticos- y notas de español (diciembre, 2023) para todos los grupos



Por otro lado, encontramos variaciones al realizar pruebas de correlación con las notas de Español y los promedios de la prueba de mecánica de lectura para ambas administraciones y según grados (ver Tabla 5). Podemos subdividir los resultados de la siguiente manera:

- **Primer grado:** sus correlaciones son moderadamente altas, oscilando entre .46 y .60 en ambas administraciones. Esto indica una relación consistente y significativa entre el desempeño en Español y las habilidades de lectura, lo que sugiere que la prueba de lectura refleja con precisión las habilidades lingüísticas que se enseñan y evalúan en las clases de Español.
- **Segundo grado:** sus correlaciones van desde .43 hasta .58. Aunque son ligeramente más bajas que en primer grado, aún representan una relación moderada entre las calificaciones de Español y las puntuaciones de las subpruebas, indicando que la relación entre las habilidades de lectura y el rendimiento en Español se mantiene.
- **Tercer grado:** sus correlaciones van de .33 a .54, lo que puede reflejar una diversificación en las habilidades académicas a medida que los(las) estudiantes maduran y el currículo se vuelve más complejo.

Tabla 5

Correlaciones de notas en Español (diciembre, 2023) con ambas administraciones de la prueba de mecánica de lectura, según grado

Grados	Sub-pruebas	Correlación con pruebas de febrero	Correlación con pruebas de mayo
Primer grado	Vocales y consonantes	.52	.52
	Sílabas	.60	.57
	Sílabas directas	.60	.57
	Sílabas inversas	.55	.58
	Variantes de sonido	.52	.58
	Grupos consonánticos	.46	.55
	Promedio de Fluidez	.48	.56
Segundo grado	Vocales y consonantes	.43	.45
	Sílabas	.53	.51
	Sílabas directas	.55	.46
	Sílabas inversas	.58	.51
	Variantes de sonido	.58	.52
	Grupos consonánticos	.57	.53
	Promedio de Fluidez	.57	.56
Tercer grado	Vocales y consonantes	.35	.33
	Sílabas	.47	.44
	Sílabas directas	.46	.39
	Sílabas inversas	.48	.42
	Variantes de sonido	.47	.42
	Grupos consonánticos	.49	.45
	Promedio de Fluidez	.54	.50

Nota. Todas las correlaciones obtuvieron un $p < .001$. Aunque descontextualizados, según estándares tradicionales (Cohen, 1988), coeficientes de correlación entre .3 a .49 tienen una magnitud media y coeficientes iguales o mayores a .05 tienen una magnitud grande.

Las implicaciones de estos resultados son varias y todas igual de importantes tanto para la evaluación de la utilidad de la prueba de mecánica de lectura, así como para las estrategias educativas en Español:

- 1) **En todos los grados, las correlaciones son positivas, lo que implica que los(las) estudiantes con mejores calificaciones en español tienden a tener puntuaciones más altas en las pruebas de mecánica de lectura. Esto confirma la validez de la prueba como una herramienta útil para medir habilidades que son relevantes para el éxito académico.**
- 2) Las correlaciones más altas en Primer grado sugieren que en las primeras etapas de la educación primaria, hay una relación más directa entre las habilidades de lectura y el rendimiento académico. Esto puede indicar que, en los primeros años, cuando los(las) estudiantes están estableciendo sus habilidades de lectura, estas habilidades son cruciales para su rendimiento en la escuela.

- 3) Las correlaciones ligeramente más bajas en Tercer grado podrían reflejar una diversificación en las habilidades y conocimientos de los(las) estudiantes. A medida que avanzan en la educación, otros factores además de sus destrezas en la mecánica de lectura podrían empezar a influir más en su rendimiento académico. De hecho, si la prueba está diseñada para que los(las) estudiantes de Tercer grado la dominen, las variaciones en el rendimiento de la prueba podrían no reflejar tan directamente las diferencias en el rendimiento académico general.
- 4) Estos resultados podrían sugerir la necesidad de estrategias de intervención diferenciadas por grado. Por ejemplo, en los grados inferiores, el énfasis en mejorar la mecánica de lectura podría tener un mayor impacto en el rendimiento académico.
- 5) La consistencia en las correlaciones entre las dos administraciones sugiere que las habilidades evaluadas por la prueba son estables y que la prueba es confiable en el tiempo.
- 6) Las correlaciones más altas en las subpruebas de sílabas y fluidez en todos los grados podrían indicar que estas áreas son particularmente relevantes para el rendimiento académico en Español.

En resumen, **los datos apoyan la validez y la confiabilidad de la prueba de mecánica de lectura como un indicador del desempeño en Español**, con algunas variaciones esperadas a medida que los estudiantes progresan a través de los grados.

Resumen de hallazgos

En la siguiente sección, abordamos una discusión sobre los hallazgos más significativos para la prueba de mecánica de lectura. Examinaremos la naturaleza descriptiva de los datos recopilados, las diferencias notables entre las administraciones de la prueba, la magnitud de los cambios reflejados por los tamaños del efecto, la consistencia de las correlaciones entre las administraciones y, finalmente, la relevante comparación de estos resultados con las calificaciones en español. Cada uno de estos puntos aporta una perspectiva valiosa sobre la validez y la confiabilidad de la prueba, evidenciando su importancia no solo como herramienta de medición, sino también como guía para la intervención pedagógica y el enriquecimiento curricular.

Hallazgo 1: Según sus descriptivos, la prueba de mecánica de lectura tiene capacidad para diferenciar las habilidades lectoras en los grados primarios

El análisis descriptivo de los datos de la prueba de mecánica de lectura revela que los(las) estudiantes muestran una amplia gama de habilidades lectoras, con variaciones que reflejan la diversidad en el dominio de la mecánica de lectura. La distribución de las puntuaciones sugiere que mientras algunos(as) estudiantes exhiben un alto grado de competencia, otros aún enfrentan desafíos significativos. Este espectro de habilidades enfatiza la necesidad de enfoques pedagógicos diferenciados, adaptados a las necesidades individuales de los estudiantes para fomentar un desarrollo equilibrado de la lectura.

Hallazgo 2: La prueba de mecánica de lectura tiene capacidad para fundamentar los cambios de promedios grupales a través del tiempo

La comparación entre las administraciones de febrero y mayo muestra mejoras en las puntuaciones, lo que indica un progreso en las habilidades de lectura a lo largo del año académico. Este avance refuerza la importancia de la instrucción continua y sistemática en la mecánica de lectura. La prueba demuestra su utilidad al medir cambios significativos en el tiempo, lo cual es esencial para evaluar la efectividad de las estrategias educativas implementadas.

Hallazgo 3: La prueba de mecánica de lectura da cuenta de la magnitud de estos cambios en promedio a través del tiempo

Los tamaños del efecto encontrados en nuestro estudio sugieren que, mientras algunas áreas de la mecánica de lectura no variaron sustancialmente, otras experimentaron cambios notables. Estos hallazgos nos instan a investigar más a fondo las prácticas educativas y el currículo, para identificar y fortalecer aquellos aspectos que contribuyen al desarrollo significativo de habilidades lectoras.

Hallazgo 4: Debido a las altas correlaciones entre administraciones, la prueba de mecánica de lectura cuenta con fiabilidad temporal, una medida importante de validez

Las altas correlaciones entre las administraciones subrayan la fiabilidad de la prueba de mecánica de lectura. El hecho de que los estudiantes mantengan rangos de desempeño similares en dos puntos distintos del año académico confirma que la prueba proporciona una medida consistente de las

habilidades de lectura. Esto es crucial para los educadores al monitorear el progreso y al realizar ajustes instruccionales informados.

Hallazgo 5: Al correlacionar con las notas de Español, la prueba de mecánica de lectura culmina estableciendo su validez, estableciendo su relevancia y utilidad en la educación primaria

La correlación positiva entre las puntuaciones de la prueba y las calificaciones de español ilustra que la mecánica de lectura es un componente integral del rendimiento académico general. La prueba, al estar alineada con los indicadores de éxito en Español, valida su relevancia práctica y su rol en la evaluación educativa integral. Este vínculo fortalece la argumentación de que la mecánica de lectura es fundamental para el aprendizaje en los grados primarios. Un instrumento como la prueba de mecánica de lectura podría ser esencial para conocer, diferenciar y poder ajustar intervenciones educativas, según el desarrollo en el desempeño de cada estudiante.

Discusión sobre las implicaciones para la prueba de mecánica de lectura

Estos hallazgos colectivos implican que la prueba de mecánica de lectura no solo es una herramienta diagnóstica eficaz, sino también un barómetro de desarrollo educativo. La eficacia de la prueba como herramienta diagnóstica se evidencia en su capacidad para identificar con precisión el rango completo de habilidades lectoras en los(las) estudiantes. A través de su implementación, hemos podido detectar tanto áreas de excelencia como de dificultad en la mecánica de lectura, permitiendo a los(las) educadores(as) enfocarse en intervenciones específicas que abordan las necesidades individuales. Esta detección temprana y precisa es crucial, ya que **la mecánica de lectura es una base sobre la cual se construyen habilidades más complejas. Por ende, una comprensión detallada de estas habilidades fundamentales permite una intervención temprana y efectiva**, lo que puede tener un impacto positivo duradero en el trayecto educativo del estudiante.

Por otro lado, la prueba de mecánica de lectura también funciona como un indicador del progreso educativo general. La correlación significativa entre las puntuaciones de la prueba y las calificaciones de Español sugiere que la prueba no solo mide las habilidades mecánicas de lectura, sino que también refleja el rendimiento académico en un área curricular clave. Esto **proporciona a los(las) educadores(as) una herramienta valiosa para monitorear y apoyar el desarrollo académico, ofreciendo la oportunidad de celebrar éxitos y de intervenir de manera personalizada donde es más necesario**. La utilidad de la prueba como reflejo del rendimiento académico subraya su papel no solo en la medición del aprendizaje, sino también en la promoción y mejora continua de la educación.

Consideraciones adicionales

Aunque la información a nivel grupal se puede utilizar para tomar ciertas decisiones, también consideramos importante nunca perder de perspectiva la necesidad de profundizar en estos resultados de manera más práctica y local. Estos resultados siempre deben ser interpretados en el contexto del programa educativo y las características específicas de los(las) estudiantes. Por ejemplo, factores como el contenido de las lecciones, las metodologías de enseñanza y el entorno del aula pueden influir en el aprendizaje. Consideremos también otros factores externos que podrían haber influido en el rendimiento de los(las) estudiantes, como cambios en el entorno de aprendizaje o en la motivación. Una prueba estandarizada nunca será sustituta de ese conocimiento y experiencia a nivel de estudiante, escuela y la comunidad que le rodea.

Conclusión

En la siguiente sección, abordaremos una serie de recomendaciones enfocadas en mejorar el manejo, análisis y divulgación de los resultados obtenidos en la prueba de mecánica de lectura. Estas recomendaciones están diseñadas para optimizar la utilidad de la prueba como herramienta educativa, garantizando que los hallazgos se analicen con precisión y se comuniquen de manera efectiva a todas las partes interesadas. Posteriormente, presentaremos sugerencias para investigaciones adicionales y posibles extensiones del estudio actual. Estas propuestas están orientadas a explorar nuevas áreas de interés, profundizar en el entendimiento de los datos existentes y expandir el alcance del estudio para abarcar contextos educativos más amplios y diversas poblaciones estudiantiles.

Recomendaciones

Ya habiendo establecido la validez, confiabilidad y, por tanto, la utilidad de la prueba de mecánica de lectura, utilizarla va a implicar un grado de compromiso para sostener estos resultados y la calidad del análisis. Para ello, y a continuación, ofrecemos un listado de recomendaciones:

Primero: Es necesario separar las administraciones de la prueba.

Para mejorar la captura de diferencias y progreso grupal en la aplicación y administración de la prueba de mecánica de lectura, es recomendable considerar una separación más amplia entre las administraciones de la prueba. Si bien esto modificaría la línea base establecida en nuestros análisis previos, un intervalo mayor permitiría una mejor observación del desarrollo de habilidades de lectura y los efectos de las intervenciones educativas a lo largo del año escolar. Esta adaptación en el cronograma de la prueba es esencial para proporcionar una visión más clara y significativa del progreso académico de los(las) estudiantes y para facilitar la implementación de estrategias pedagógicas más efectivas y oportunas. Idealmente, la prueba se administraría en tres ocasiones: al comienzo del año escolar, a mitad y al final; la siguiente mejor opción sería al comienzo y al final del año escolar.

Segundo: Debemos evaluar y analizar estos resultados consistentemente.

Es fundamental que la evaluación y análisis de los resultados de la prueba de mecánica de lectura se realicen de manera continua y sistemática. Mantener una metodología coherente y consistente en la aplicación y revisión de la prueba es clave para preservar la integridad de los análisis longitudinales. Dicha consistencia asegura que las variaciones observadas en el rendimiento de los(las) estudiantes reflejen cambios genuinos en su aprendizaje y no sean artefactos de modificaciones metodológicas. Este enfoque meticuloso es esencial para la toma de decisiones informadas sobre intervenciones pedagógicas individualizadas. Un seguimiento riguroso permite identificar tendencias, medir el impacto de las estrategias educativas implementadas y ajustar los planes de estudio para atender las necesidades educativas de cada estudiante de manera efectiva y oportuna.

Tercero: Al establecer estos resultados como línea base, debemos definir los casos extremos.

La definición precisa de lo que constituye un caso extremo en el contexto de las pruebas de mecánica de lectura es crucial, especialmente en un entorno educativo tan variado como el de primer grado. Aunque los cortes realizados en este estudio se basaron en análisis estadísticos, la experiencia educativa y el conocimiento del contexto específico de aprendizaje deben guiar la identificación y el tratamiento de estos casos. Esto es esencial porque lo que se considera un caso extremo puede variar significativamente dependiendo de factores como el nivel de desarrollo del(de la) estudiante, el currículo y el entorno educativo. Una comprensión integral del contexto educativo permitirá una interpretación más precisa de estos casos y ayudará a tomar decisiones más informadas sobre intervenciones y adaptaciones curriculares. Al definir los casos extremos no solo por criterios estadísticos, sino también por consideraciones pedagógicas, podemos asegurarnos de que el análisis de la prueba refleje con mayor fidelidad las realidades y necesidades de los(las) estudiantes.

Cuarto: Dar seguimiento a estos estudios para fortalecer la literacidad de datos y generar una actitud de la educación basada en evidencia.

Dar seguimiento a estudios como el de la prueba de mecánica de lectura es fundamental para fomentar una cultura de literacidad de datos en el ámbito educativo. Al analizar y utilizar los datos de manera continua y reflexiva, se promueve una actitud de educación basada en la evidencia, donde las decisiones pedagógicas y curriculares se toman con base en hallazgos concretos y no en suposiciones. Este enfoque no solo mejora la calidad de la enseñanza y el aprendizaje, sino que también empodera a educadores(as) y administradores(as) para adaptarse de manera más efectiva a las necesidades de los(las) estudiantes. La práctica de analizar datos educativos de manera rigurosa y sistemática contribuye a una mayor comprensión y eficiencia en el proceso educativo, asegurando que cada decisión tomada sirva para enriquecer y mejorar la experiencia de aprendizaje.

Quinto: Es bien merecido divulgar y celebrar estos resultados.

Es imprescindible divulgar y celebrar los resultados obtenidos en la prueba de mecánica de lectura, ya que demuestran de manera convincente el impacto positivo y tangible de la educación en el aprendizaje de los(las) estudiantes. Los avances significativos observados en las habilidades de lectura son un claro testimonio del arduo trabajo y la dedicación de educadores(as) y alumnos(as) por igual. Estos hallazgos refuerzan la idea de que la educación es una herramienta poderosa para el desarrollo intelectual y personal, y que las estrategias de enseñanza implementadas están facilitando un aprendizaje efectivo. Celebrar estos éxitos no solo valida los esfuerzos de todos(as) los(as) involucrados(as) en el proceso educativo, sino que también sirve como un recordatorio inspirador de que el compromiso con la educación de calidad tiene un impacto profundo y duradero en la vida de los(las) estudiantes.

Pasos futuros

Los pasos futuros para el estudio de la prueba de mecánica de lectura se orientan hacia la expansión y profundización del análisis para mejorar continuamente la calidad educativa. En primer lugar, es esencial dar seguimiento al análisis longitudinal ya iniciado. Esta continuación permitirá no solo ver el progreso a lo largo del tiempo, sino también identificar patrones y tendencias que pueden influir en la toma de

decisiones pedagógicas. Este análisis longitudinal es vital para entender cómo las habilidades de lectura evolucionan y cómo las intervenciones educativas impactan en el aprendizaje a largo plazo.

Además, es importante realizar comparaciones más extensas entre diferentes regiones educativas. Esta expansión del estudio permitirá una comprensión más completa de las variaciones en el rendimiento de lectura, posiblemente debido a diferencias en los recursos, metodologías de enseñanza y contextos socioeconómicos. Incluir variables adicionales, como el estatus socioeconómico y la pertenencia a programas de educación especial, enriquecerá aún más el análisis, ofreciendo una visión más detallada y completa de los factores que influyen en la mecánica de lectura. Estos datos más completos y detallados serán fundamentales para diseñar intervenciones más efectivas y personalizadas.

Finalmente, publicar los resultados en foros y revistas académicas, tanto a nivel nacional como internacional, contribuirá a establecer comparativas y validar nuestros hallazgos en un contexto más amplio. Esto no solo aumentará la visibilidad del estudio, sino que también facilitará el intercambio de conocimientos y experiencias con otros(as) profesionales en el campo educativo. Paralelamente, es crucial establecer una relación de mentoría con el Departamento de Educación. Esta colaboración permitirá transferir conocimientos y habilidades analíticas, asegurando que el Departamento pueda realizar análisis similares de forma independiente en el futuro. Este enfoque colaborativo y de empoderamiento es clave para un desarrollo educativo sostenible y basado en la evidencia.

Cierre

En resumen, nuestro estudio exhaustivo de la prueba de mecánica de lectura ha revelado hallazgos significativos sobre el desarrollo de habilidades lectoras en estudiantes de grados primarios. Hemos identificado mejoras en el rendimiento a lo largo del tiempo, establecido correlaciones importantes entre la mecánica de lectura y el rendimiento académico en español, y confirmado la validez y fiabilidad de la prueba como herramienta de evaluación. Estos resultados subrayan la importancia de una enseñanza de lectura eficaz y basada en datos, resaltando la necesidad de seguir investigando y adaptando nuestras prácticas educativas. A medida que avanzamos, la continua evaluación, análisis y publicación de estos hallazgos no solo enriquecerán nuestro entendimiento del aprendizaje lector, sino que también contribuirán al avance de una educación de calidad, inclusiva y efectiva para todos los(las) estudiantes.

Referencias

- Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114(3), 494.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
- DeWilde, V., Brysbaert, M. y Eyckmans, J. (2022). Formal versus informal L2 learning: How do individual differences and word-related variables influence French and English L2 vocabulary learning in Dutch-speaking children? *Studies in Second Language Acquisition*, 44, 87-111.
- McPadden, D. y Brewe, E. (2017). Impact of the second semester University Modeling Instruction course on students' representation choices. *Physical Review Physics Education Research*, 13, 020129.
- Nielsen, K. (2023). Why can the flipped classroom frustrate students? Experiences from an Engineering Mathematics course. *Education Sciences*, 13(4), 396.
- Palmgren, E., Tuominen, K. y Kontro, I. (2022). Self-efficacy and conceptual knowledge in quantum mechanics during teaching reforms and the COVID-19 pandemic. *Physical Review Physics Education Research*, 020122.