

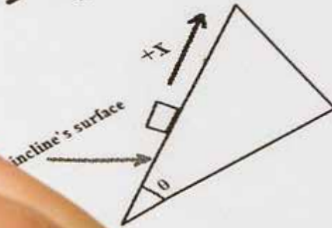
around the Earth with a radius $r_1 = R$.
 around the Earth with a radius $r_2 = 2R$.
 of a satellite and a represent the magnitude of a satellite's acceleration. Which
 gives the correct relation between the speeds and accelerations of the satellites?

- (D) $v_2 = \frac{1}{2}v_1$; $a_2 = \frac{1}{2}a_1$
 (E) $v_2 = v_1$; $a_2 = \frac{1}{2}a_1$



with constant speed around a horseshoe-shaped path as shown with
 in the figure. Which one of the following choices best describes the
 of the average acceleration of the car in traveling from W to X?

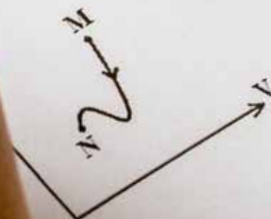
- (B) ← (C) ↗ (D) → (E) There is no average acceleration



A mass on a frictionless incline has a gravitational force
 the incline, and a force applied by a person that
 The mass remains at rest and the incline makes an angle
 from the horizontal. Which one of the following
 orientations of the applied force best describes the
 upward, parallel to the incline's surface?

- (A) The applied force is perpendicular to the incline.
 (B) The applied force is parallel to the incline and points up.
 (C) The applied force is parallel to the incline and points down.
 (D) The applied force is perpendicular to the incline and points down.
 (E) This is a completely impossible situation.

axis.
 Which one of the
 the total
 for



Efecto de un programa de entrenamiento basado en las pruebas Saber 11 de matemáticas

Effect of a training program based on the Saber 11 test in mathematics

Henry Daza Chaves

Docente, Universidad Nacional Abierta y a Distancia – UNAD, Pasto, Colombia, ORCID: 0000-0001-9814-7702

Jonathan Andrés Estrada

Estudiante de licenciatura en Matemáticas, Universidad Nacional Abierta y a Distancia – UNAD, Pasto, Colombia, ORCID: 0000-0002-4215-8952

María Camila Guerrero

Estudiante de licenciatura en Matemáticas, Universidad Nacional Abierta y a Distancia – UNAD, Pasto, Colombia, ORCID: 0000-0003-0141-7345

RESUMEN

Al finalizar la educación media en Colombia, los estudiantes deben presentar las pruebas saber 11, las cuales evalúan cinco competencias: matemáticas, sociales y ciudadanas, lectura crítica, inglés y ciencias naturales. Independientemente de una larga trayectoria de formación de alrededor de 11 años, los estudiantes suelen tomar programas de entrenamiento que los preparen para estas pruebas, con la expectativa de que esto pueda incidir en su desempeño. El objetivo de este estudio fue evaluar el efecto que un programa de entrenamiento ejerce sobre el desempeño de los estudiantes para dar respuesta a problemas y situaciones basados en la prueba Saber 11 de matemáticas. Se empleó un diseño preexperimental de preprueba y posprueba y se aplicó un conjunto de simulaciones retroalimentadas a un grupo de 46 estudiantes de grados décimo y undécimo de dos instituciones educativas de Nariño-Colombia, a través de un ambiente virtual de aprendizaje. Basados en la medición interna, los resultados mostraron un incremento del 14 % en el desempeño de los estudiantes después del entrenamiento. El programa, aparentemente, fue altamente efectivo (0,95) según la métrica de Cohen. Aunque los resultados del entrenamiento fueron satisfactorios, se reconoce que es un estudio sin grupo de control y que, por tanto, es recomendable contrastar los hallazgos mediante un estudio experimental puro.

PALABRAS CLAVE:

competencia, desempeño, entrenamiento, prueba.

ABSTRACT

At the end of secondary education in Colombia, students must take the Saber 11 test, which evaluates five competencies: mathematics, social and civic skills, critical reading, English and natural sciences. Regardless of a long training path of around eleven years, students often take training programs that prepare them for this test, with the expectation that this may affect their performance. The objective of this study was to evaluate the effect that a training program has on the performance of students to respond to problems and situations based on the Saber 11 test in mathematics. A pre-test and post-test pre-experimental design was used, and a set of feedback simulations was applied to a group of 46 tenth and eleventh grade students from two educational institutions in Nariño-Colombia, through a virtual learning environment. Based on internal measurement, the results showed a 14% increase in student performance after training. The program was apparently highly effective (0.95) by Cohen's metric. Although the results of the training were satisfactory, it is recognized that it is a study without a control group and that, therefore, it is advisable to contrast the findings through a pure experimental study.

KEYWORDS:

Competence, performance, test, training.

INTRODUCCIÓN

Las pruebas Saber 11 son uno de los diferentes instrumentos que emplea el Ministerio de Educación Nacional para monitorear la calidad de la educación secundaria en Colombia. Estas pruebas tienen las funciones de evaluar por medio de instrumentos estandarizados a los estudiantes de grado 11° en las cinco competencias que se mencionaron en el resumen y, por otro lado, recoger información sobre un conjunto de atributos, principalmente socioeconómicos y demográficos, que son utilizados para construir modelos estadísticos e investigar factores que estén incidiendo en el desarrollo de dichas competencias en los evaluados. Estas investigaciones, tanto internas (al interior del ICFES) como externas (investigadores particulares), aportan información para la toma de decisiones en materia educativa.

Uno de los principales objetivos de la prueba Saber 11, según el Instituto Colombiano para la Evaluación de la Educación ICFES (2022a, p. 11) es comprobar “el grado de desarrollo de las competencias de estudiantes que están por finalizar grado undécimo de la educación media”. Los resultados de estas pruebas aportan información al estudiante para que pueda orientar su proyecto de vida, así como también aportan información a los colegios para la toma de decisiones en materia curricular sobre sus proyectos educativos institucionales con fundamento en los estándares básicos de competencias. También, son usadas por algunas instituciones de educación superior como una fuente de información para clasificar a los estudiantes que acceden a sus cupos universitarios.

De acuerdo con el informe nacional presentado por el ICFES (2022b, p.24), “entre 2014 y 2021 se presentó un descenso del promedio del puntaje global” con respecto al calendario A, lo cual específicamente sig-

nificó que se pasara de 255 puntos promedio a 250, representando una gran disminución “de acuerdo con el tamaño del efecto” (ICFES, 2022b, p.24). En los gráficos de la página 25 de este informe se resalta que, entre 2017 y 2021, en promedio, alrededor de un 8% de los evaluados solamente logra uno de los cuatro niveles de desempeño en la competencia de matemáticas y alrededor de un 38% alcanzan los niveles 1 y 2. Estos resultados que aplican para el calendario A, cuya población representa alrededor de un 97% de los examinados, indicarían que cerca de la mitad no logran los niveles 3 y 4, lo que significa que no evidencian la capacidad para “resolver problemas y justificar la veracidad o falsedad de afirmaciones que requieren el uso de conceptos de probabilidad, propiedades algebraicas, relaciones trigonométricas y características de funciones reales. Lo anterior, en contextos principalmente matemáticos o científicos abstractos [nivel 3]” (ICFES 2020, p. 3); en consecuencia, tampoco reflejan la capacidad para “seleccionar información, señalar errores y hacer distintos tipos de transformaciones y manipulaciones aritméticas y algebraicas sencillas; esto para enfrentarse a problemas que involucren el uso de conceptos de proporcionalidad, factores de conversión, áreas y desarrollos planos, en contextos laborales u ocupacionales, matemáticos o científicos, y comunitarios o sociales [nivel 4]” (ICFES 2020, p. 4).

Teniendo en cuenta que durante los últimos años los resultados obtenidos en la prueba Saber 11 a nivel nacional no han suplido las expectativas del Ministerio de Educación Nacional sobre los referentes de calidad, surge la necesidad de implementar planes y estrategias a corto, mediano y largo plazo que propendan por el mejoramiento del desarrollo de las competencias y redunden en un mayor rendimiento de los estudiantes en estos exámenes.

Independientemente de que son diferentes los factores que afectan la calidad de la educación, las instituciones educativas realizan, desde sus posibilidades, sus mejores esfuerzos en la trayectoria formativa de sus alumnos para desarrollar de la mejor forma sus competencias y, así, puedan obtener buenos resultados en estas pruebas; tal es el caso que cuando están próximos a presentarlas, optan por programas de entrenamiento para potencializar sus habilidades para resolver los problemas y/o situaciones potenciales de este examen. Es por tal razón que resulta importante cuestionarse sobre si entrenar a los estudiantes para este tipo de pruebas puede favorecer su desempeño en las mismas, pues partiendo de una analogía con el deporte, la música u otras disciplinas, el entrenamiento siempre ha sido un factor clave para optimizar el desempeño y/o rendimiento en dichas prácticas.

Este artículo presenta los principales resultados de un proyecto aplicado desarrollado por estudiantes de Lic. en Matemáticas de la Universidad Nacional Abierta y a Distancia (UNAD), el cual consistió en aplicar un programa de entrenamiento basado en pruebas Saber 11 de Matemáticas y analizar su efecto. El proceso se desarrolló con estudiantes de grados 10° y 11° de dos establecimientos educativos de Nariño (Colombia) mediante un ambiente virtual de aprendizaje. Los detalles de dicho proyecto se pueden consultar con en Estrada y Guerrero (2022).

MARCO TEÓRICO

El entrenamiento se define, según Meza (2012, p. 116), como la “adquisición de habilidades, capacidades y conocimientos como resultado de la exposición a la enseñanza de algún tipo de oficio, carrera o para el desarrollo de alguna aptitud física o mental y que está orientada a reportarle algún beneficio o utilidad al individuo que se somete a tal

o cual aprendizaje”. Esto justifica el hecho de que los grupos musicales, los grupos de baile, los grupos deportivos, o en particular, una persona que practique alguna actividad, tenga que desarrollar procesos de entrenamiento constante para optimizar su desempeño y/o rendimiento. Lo mismo ocurre cuando un profesor programa un examen sobre un tema determinado; los estudiantes se preparan para ese momento ejercitando su conocimiento mediante la lectura de los temas abordados y anticipando respuestas a preguntas potenciales de dicho examen.

Un estudio realizado por Díaz et al. (2017, p. 99), encontró que, por ejemplo, entrenar a los estudiantes en autorregulación, en efecto, mejora su autorregulación.

Por otro lado, investigadores de la Universidad de Cádiz aplicaron un programa de entrenamiento computarizado para mejorar el aprendizaje de matemáticas en niños con riesgo de presentar dificultades de aprendizaje, el cual mostró que “el logro matemático de los estudiantes pertenecientes al grupo experimental mejoró su media en 11,67 puntos, siendo superior a la ganancia obtenida por el grupo control, que incrementó su media en 1,13 puntos” (Guzmán et al., 2015, p. 112).

Una investigación que trata sobre la transferencia de un entrenamiento de memoria de trabajo (MT) a las habilidades académicas y estrategias de resolución de problemas al inicio de la escolaridad mostró “efectos que se transfieren al aprendizaje, en particular en el rendimiento de tareas matemáticas y, en menor medida, en lengua” (López y Arán, 2021, p. 116).

En el sentido que la resolución de problemas matemáticos implica la comprensión lectora, es importante destacar estudios como el de Campos y Talavera (2014, p. 13), en el que, al realizar entrenamientos con base en esta

estrategia, encontraron beneficios significativos sobre la solución de problemas aritméticos de enunciado verbal.

Un aspecto importante al desarrollar estos programas de entrenamiento para fortalecer determinadas habilidades es que los estudiantes juegan un papel activo en el proceso, lo que implica que, según el cono de aprendizaje de Edgar Dale, simular experiencias reales permite recordar en un corto plazo un 90% de lo que se dice o se hace (Arredondo et al., 2010, p. 67).

METODOLOGÍA

Para el programa de entrenamiento empleado en este estudio se elaboraron preguntas de selección múltiple con única respuesta basadas en contenidos de estadística, geometría, álgebra y cálculo, construidas sobre competencias como: argumentación, ejecución, formulación, representación e interpretación, teniendo como referencia los estándares básicos de competencias del Ministerio de Educación Nacional. Lo anterior, entendiendo que es “necesario enseñar a los estudiantes las habilidades con diferentes tipos de contenido educativo para que entiendan cómo aplicarlas” (Schunk, 2010, p. 75).

Se plantearon nueve simulacros, cada uno con 10 preguntas que se seleccionaban de manera aleatoria de un repositorio. Adicionalmente, cada simulacro contaba con tres intentos, lo que constituía la posibilidad de que cada individuo pueda realizar 27 ensayos. Es necesario destacar que el incremento de ensayos se relaciona inversamente con el promedio de errores según la curva de aprendizaje propuesta por Schunk (2010, p. 182).

ENFOQUE

Se emplea el enfoque cuantitativo, el cual cuenta con un conjunto de características

que no se detallarán en este artículo, pero es un enfoque que se define básicamente como un proceso sistemático que “utiliza la recolección de datos para probar hipótesis con base en la medición numérica y el análisis estadístico, con el fin de establecer pautas de comportamiento y probar teorías” (Sampieri et al., 2014, p. 4).

POBLACIÓN Y MUESTRA

El grupo poblacional estuvo compuesto por alumnos de grados 10° y 11° de dos establecimientos educativos del departamento de Nariño, uno ubicado en el sector rural del municipio de La Cruz y el otro del sector urbano de la ciudad de Pasto. En total participaron 59 estudiantes del programa de entrenamiento, de los cuales para el análisis se conformó un grupo de 46 individuos que realizaron completamente el proceso; es decir, participaron tanto en la preprueba, como en el desarrollo del proceso de entrenamiento y la posprueba.

No se emplea muestreo probabilístico, dado que el grupo poblacional fue pequeño y el tamaño de la muestra fue de 46 individuos, cercano al tamaño al de la población objetivo. El fin del estudio no pretendía extrapolar los resultados a una gran población, sino estudiar, de forma preexperimental (sin grupo de control), el efecto que un programa de entrenamiento puede producir sobre el desempeño de un pequeño grupo de estudiantes para resolver problemas y situaciones basados en el examen oficial Saber 11 sobre la competencia de matemáticas. La idea es que esta experiencia pueda orientar, a futuro, un estudio experimental puro de mayor complejidad que incluya las distintas competencias y que pueda validarse a través de los resultados oficiales de las pruebas Saber 11.

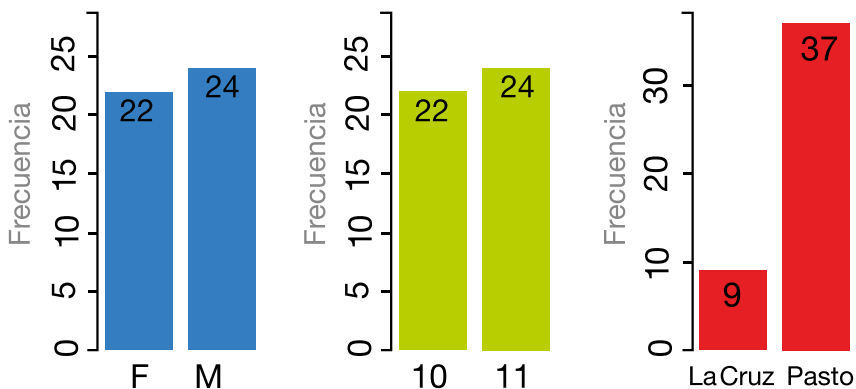
DISEÑO

Como se ha mencionado anteriormente, el diseño es de tipo preexperimental (no hay grupo de control). Este diseño consiste en que “a un grupo se le aplica una prueba previa al estímulo o tratamiento experimental; después se le administra el tratamiento y finalmente se le aplica una prueba posterior al estímulo” (Sampieri, et al., 2014, p. 141). Se caracteriza por su bajo grado de control de los factores y fuentes que puedan invalidar los resultados, no obstante, puede utilizarse como una aproximación exploratoria al fenómeno y servir de soporte para diseñar un experimento puro. Se resalta también que las condiciones sobre las cuales se desarrolló el proyecto aplicado, particularmente las de las instituciones educativas, no permitan la experimentación pura, pues el proyecto de alguna u otra manera buscaba llevar un beneficio que garantice el acceso a todos los estudiantes focalizados, por lo cual, crear un grupo de control obstaculizaba ese objetivo.

Análisis descriptivo

Los gráficos de la figura 1 indican que, por grado y por género los grupos estuvieron equilibrados, mientras que, por ubicación, la institución educativa de Pasto fue la que tuvo la mayor cantidad de estudiantes en el programa de entrenamiento.

Figura 1. Número de estudiantes por ubicación, grado y género.



Nota. Las frecuencias son absolutas.

INSTRUMENTOS

Se emplearon dos cuestionarios compuestos por 25 preguntas cada uno. Uno de ellos es aplicado en la preprueba y el otro en la posprueba. Para el entrenamiento se aplicaron nueve simulacros, cada uno compuesto por 10 preguntas seleccionadas y ordenadas aleatoriamente desde un banco de 135 preguntas. Los problemas y situaciones planteadas, tanto en los simulacros como en las pruebas, fueron de selección múltiple con única respuesta. Para profundizar sobre el **control de validez y confiabilidad**, así como el análisis de los **índices de dificultad** de los instrumentos, se puede consultar la publicación de Estrada y Guerrero (2022, pp. 45-57), cuyo enlace se encuentra en la lista de referencias.

RESULTADOS

Los siguientes datos analizados fueron tomados los anexos de la publicación de Estrada & Guerrero (2022).

La tabla 1 registra los promedios obtenidos por los estudiantes tanto en la preprueba como en la posprueba. Aunque se deja claro que el presente estudio no busca una generalización de los hallazgos a una gran población, más adelante se hará una verificación inferencial de la significancia de la diferencia media de 3,5 puntos que se produjo.

Tabla 1. Medias y diferencias de medias para la preprueba y la posprueba.

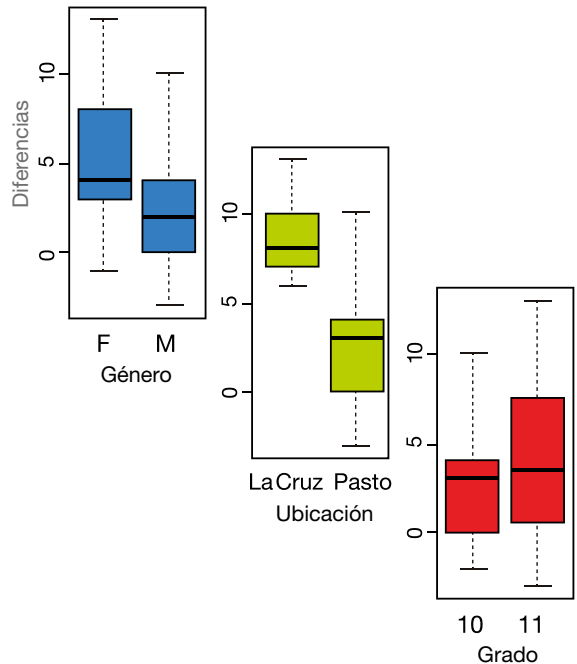
	Valor	Estudiantes evaluados
Preprueba	7,56 puntos promedio sobre 25.	46
Posprueba	11,06 puntos promedio sobre 25.	46
Diferencia:	3,5 puntos promedio sobre 25.	

Nota: Los puntos promedio tanto para la preprueba como la posprueba se obtuvieron al sumar los aciertos por individuo y dividir entre el total de individuos (46). La diferencia de medias se obtiene al restar el puntaje medio de la posprueba con el puntaje medio de la preprueba.

Los gráficos de la figura 2 permiten apreciar una diferencia en la tendencia de los resultados por ubicación y por género. El gráfico de la izquierda y el del centro parecen mostrar que el programa fue más efectivo para el género femenino y para estudiantes de la Cruz (N.), aunque se probará más adelante si existe una diferencia estadísticamente

significativa al igual que para el grado. Aunque las diferencias para algunos grupos muestran cambios en la dispersión, no se aprecian datos atípicos.

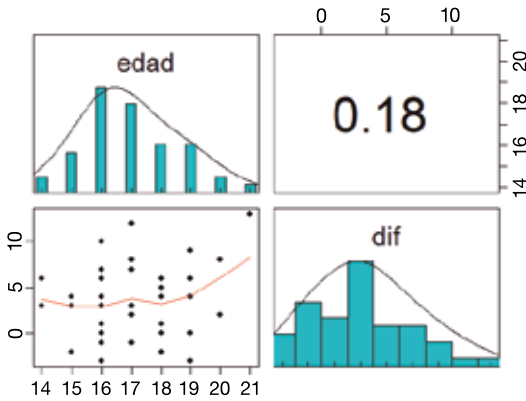
Figura 2. Diferencias en el desempeño por ubicación, grado y género.



Nota. El eje “Y”, denominado diferencias, es una variable adicional creada partir de restar los resultados individuales de la posprueba y la preprueba. Posteriormente esta nueva variable se compara en función de variables que caracterizan a los individuos.

El valor de la correlación (0,18) de la figura 3 indica una posible relación directa entre la edad y las diferencias; es decir que, posiblemente, a mayor edad, los estudiantes pudieron presentar tener un mayor rendimiento. Esto se probará de forma inferencial más adelante. La curva de densidad de las diferencias parece distribuir normalmente.

Figura 3. Gráfico de correlación múltiple entre las diferencias y la edad del estudiante.



Nota. El gráfico muestra la densidad de probabilidad de la edad y las diferencias; la dispersión entre estas variables; y, su respectiva correlación.

Análisis inferencial

Como se mencionó anteriormente, el propósito no es generalizar los hallazgos a una gran población. No obstante, se aporta un análisis desde una perspectiva inferencial que permita complementar los resultados indicados en la parte descriptiva. Se inicia por la aplicación de la prueba T-Student para determinar si la diferencia media de 3,5 puntos sobre 25 observada en el análisis descriptivo es estadísticamente significativa.

Partimos de la formulación Mendenhall et al. (2010 p. 412), cuyo estadístico T viene dado por:

$$T = \frac{\bar{d}}{S_d / \sqrt{n}}$$

Donde

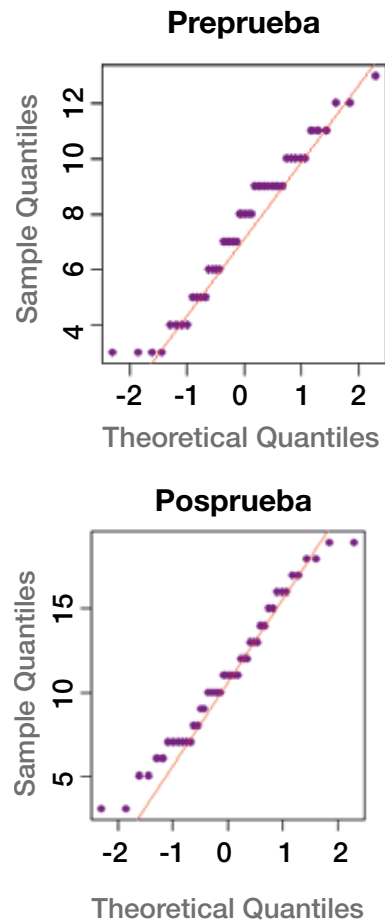
n : Número de diferencias pareadas.

\bar{d} : Media de las diferencias muestrales.

S_d : Desviación estándar de las diferencias muestrales.”

La aplicación de este estadístico implica que los datos en las pruebas aplicadas distribuyan normalmente (Caughlin, 2022). Un acercamiento inicial por medio de los gráficos de la figura 4 permite apreciar aparentes problemas de normalidad tanto en la preprueba como en la posprueba, dada la forma no ajustada a la línea teórica de algunos puntos en los extremos de las distribuciones.

Figura 4. Gráficos de normalidad de las pruebas aplicadas.



Nota. Los gráficos muestran qué tanto se ajustan los datos (puntos) a la distribución teórica (línea).

Para verificar formalmente la normalidad de los datos se recurre a la prueba de Shapiro Wilk, cuya formulación no se relacionará en este artículo por asuntos de espacio, pero que puede consultarse en la publicación de Tapia y Ceballos (2021, pp. 87-88). Cabe mencionar que esta prueba de normalidad se aplica para muestras menores a 50 observaciones (Tapia y Ceballos, 2021), como es este el caso. Los resultados presentados en la tabla 2 indican que, tanto la preprueba como la posprueba, distribuyen normalmente a un 0,05 de nivel de significancia. Esto implica viabilidad para la aplicación del estadístico T-Student.

Nota. H₀ indica hipótesis nula y H_a indica hipótesis alterna.

Tabla 2. Supuestos de normalidad de los datos.

	Preprueba	Posprueba
W	0,95952	0,97151
P-Valor	0,1098	0,3149
H ₀ : Los datos provienen de una población normal.	Cumple	Cumple
H _a : Los datos NO provienen de una población normal.	No cumple	No cumple
Nivel de significancia	0,05	0,05

De acuerdo con el P-valor indicado en la tabla 3, se aprecia una diferencia estadísticamente significativa de 3,5 puntos promedio sobre 25 en las pruebas aplicadas, indicando que los estudiantes tuvieron un incremento promedio del 14 % en los resultados, lo cual puede estar vinculado al programa de entrenamiento que recibieron a través del ambiente virtual de aprendizaje.

Tabla 3. Resultados de la aplicación de la prueba T-Student.

Prueba T pareada			
Valor T	6,0644	Grados de libertad	45
Diferencia media sobre 25.	3,5	P-valor	0,0000
Hipótesis nula	$H_0: \overline{X_D} = 0$	Hipótesis alternativa	$H_1: \overline{X_D} \neq 0$
Significancia	0,05		

Nota. $\overline{X_D}$ Representa la media del vector de diferencias de las pruebas aplicadas (postest – pretest).

Análisis inferencial complementario por género, ubicación, grado y edad.

Se realizó un modelo ANOVA de un solo factor, cuyo p-valor mostrado en la tabla 4 indica que existen diferencias estadísticamente significativas en el desempeño por género a un nivel de significancia del 0,05. El objetivo del análisis de varianza es determinar diferencias significativas de medias muestrales (Walpole, et al., 2010).

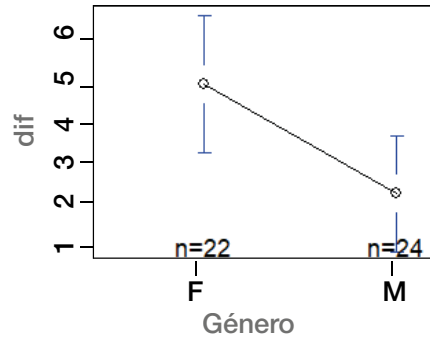
Tabla 4. Modelo ANOVA para evaluar diferencia de medias por género M o F.

Efecto	g.l.	SC.	CM.	F	P-Valor
Género	1	89,2	89,21	6,539	0,0141
Residuales	44	600,3	13,64		

Nota. La hipótesis nula para estos modelos afirma que no existen diferencias significativas en los grupos con un p-valor igual o mayor al 0,05. Convecciones: g.l: grados de libertad, SC: suma de cuadrados, CM: cuadrados medios, F: valor de la distribución F-Snedecor.

El gráfico de la figura 5 permite apreciar un posible efecto del proceso de entrenamiento significativamente mayor para el género femenino.

Figura 5. Gráfico de medias por género con respecto a las diferencias.



Nota. El eje “dif” corresponde al vector de datos resultante de restar los vectores post-test y pretest. Los valores de n=22 y n=24, indican la cantidad de individuos por género en la muestra.

Para verificar puntualmente la diferencia, se recurre a la comparación múltiple de Tukey, cuyo método es “más conservador para comparar pares de medias de tratamientos” (Gutiérrez y Salazar, 2008). Los resultados mostrados en la tabla 5 reflejan una diferencia media del género femenino contra el género masculino de 2,79 puntos sobre 25.

Tabla 5. Comparaciones múltiples por género.

Prueba Tukey				
	Diferencia	IC-L	IC-U	P-Valor
M-F	-2,787879	-4,985079	-0,5906784	0,0140771

Nota: IC-L e IC-U indican los límites inferior y superior del intervalo de confianza para la diferencia.

Los modelos ANOVA generalmente se validan mediante supuestos como: la normalidad en los residuales, para la cual se puede emplear la prueba de Shapiro Wilk, cuyo teórico fue referido anteriormente; la homogeneidad de varianza, para la cual se puede emplear el test de Barlett, cuya formulación se puede consultar en la publicación de Correa y Rojas (2006 pp. 58-59); y la autocorrelación de los residuales mediante las pruebas de Durbin Watson o Breusch-Godfrey, cuya formulación se puede consultar en Quintana y Mendoza (2016, p.159).

Los resultados presentados en la tabla 6 indican que la diferencia presentada en la tabla 5 es estadísticamente válida.

Tabla 6. Validación de supuestos del modelo ANOVA por género.

Validación del modelo		
Test de Shapiro Wilk	Test de Bartlett	Test de Breusch-Godfrey
W = 0.96343, p-value = 0.1558	Bartlett's K-squared = 0.31795 df = 1 p-value = 0.5728	LM test = 2.7029 df = 1 p-value = 0.1002

Nota: Los resultados se validan asumiendo que la hipótesis nula se cumple para un p-valor mayor al 0.05.

Con respecto al desempeño por ubicación del establecimiento educativo (La Cruz y Pasto) y, de acuerdo con el p-valor de la tabla 7, se aprecia que efectivamente existen diferencias significativas.

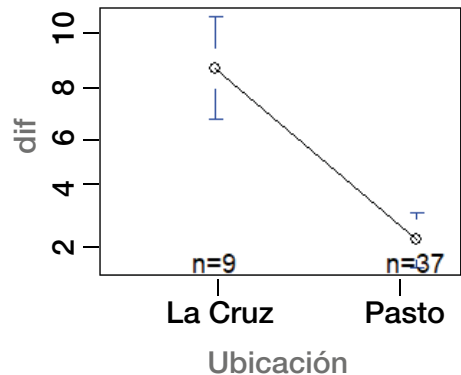
Tabla 7. Modelo ANOVA de las diferencias por ubicación.

Efecto	g.l.	SC	CM	F	P-Valor
Ubicación	1	298,7	298,69	33,63	0,0000
Residuales	44	390,8	8,88		

Nota. Los elementos de esta tabla se interpretan de forma similar al anterior modelo ANOVA.

El gráfico de la figura 6 indica un posible efecto del proceso de entrenamiento significativamente mayor para estudiantes de La Cruz.

Figura 6. Gráfico de medias por ubicación con respecto a las diferencias.



Nota. La lectura de los elementos del gráfico se realiza de forma similar a la figura 5.

La diferencia media según la prueba de Tukey indicada en la Tabla 8 es aproximadamente de 6.42 puntos sobre 25 de los estudiantes de la Cruz con respecto a los estudiantes de Pasto.

Tabla 8. Comparaciones múltiples por ubicación

Prueba Tukey				
	Diferencia	IC-L	IC-U	P-Valor
Pasto – La Cruz	-6,423423	-8,655803	-4,191044	0,0000

Nota: Los elementos de la tabla se interpretan similarmente a la anterior prueba de Tukey.

Los p-valores de la tabla 9 indicarían que el valor de la diferencia media presentado en la tabla 8 es estadísticamente válido.

Tabla 9. Validación del modelo de diferencias por ubicación.

Validación del modelo		
Test de Shapiro Wilk	Test de Bartlett	Test de Durbin Watson
W = 0,97344 p-valor = 0,3695	Bartlett's K-squared = 0,49431 df = 1 p-valor = 0,482	DW = 1,8473 p-value = 0,2517

Nota: Los elementos de la tabla se interpretan similarmente a la anterior validación.

Con respecto a las diferencias por grado, el p-valor indicado en la tabla 10, permite concluir a un 0,05 de nivel de significancia que no existen diferencias estadísticamente significativas por grado. Es decir, estar cursando grado décimo u undécimo no incidió en los posibles efectos del programa de entrenamiento.

Tabla 10. Modelo ANOVA de las diferencias por grado.

Efecto	g.l.	SC.	CM.	F	P-Valor
Grado	1	17,1	17,08	1,117	0,296
Residuales	44	672,4	15,28		

Nota. Los elementos de esta tabla se interpretan de forma similar al anterior modelo ANOVA.

Por último, se aplicó la prueba de correlación de Pearson para determinar la relación entre la edad y el rendimiento de los estudiantes en el programa de entrenamiento. Este coeficiente es calculado a partir de un procedimiento complejo que por la extensión no se detalla en este artículo; no obstante, puede consultar su formulación y un ejemplo en Mendenhall et al. (2010, p. 534).

En la tabla 11 se aprecia una correlación de 0,1829, con un p-valor que indica que no es estadísticamente significativa. Es decir, la edad no estuvo asociada con un mayor o menor desempeño del estudiante en el proceso.

Tabla 11. Prueba de correlación de Pearson entre la edad y la diferencia de las pruebas aplicadas.

Test de correlación	
Valor prueba T-Student	1,2341
P-valor	0,2237
Hipótesis nula	La correlación es igual a cero.
Nivel de significancia	0,05
Intervalo de confianza	L= -0,113 U= 0,450
Valor de correlación	0,1829

Nota. El intervalo de confianza incluye el cero (0), lo que indica que probabilísticamente dicha correlación puede estar ubicada en cero (0).

Tamaño del efecto

Se concluye la sección de los resultados con la medición del tamaño del efecto. Este permite determinar la magnitud de la diferencia media que se presentó en los resultados de la prueba T-Student (tabla 3). Esta métrica se calculó mediante la *d* - de Cohen, cuya formulación adaptada de Caughlin (2022) es:

$$d = \frac{\bar{X}_D}{S_D} \quad \text{con} \quad S_D = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

Donde \bar{X}_D representa la media de las diferencias, n_1, n_2, s_1^2 y s_2^2 son los tamaños de los grupos (*n*) y las varianzas (*s*) para la preprueba y la posprueba respectivamente.

La categorización de los valores *d* para los cuales se establece si un efecto es pequeño,

moderado o grande depende del investigador. Caycho, et al. (2016, p. 459) señala que “valores de la *d* de Cohen inferiores a 0,20 señalan la no existencia de efecto; valores entre 0,21 a 0,49 hacen referencia a un pequeño efecto; así mismo, valores oscilantes entre 0,50 a 0,70 indican un moderado efecto; finalmente, valores mayores a 0,80 señalan un efecto grande”.

Los resultados presentados en la tabla 12 indican una *d* estimada de 0,95, la cual, partiendo de Caycho, et al. (2016), señala un gran efecto. En síntesis, los resultados muestran que el entrenamiento puede ser altamente efectivo en el rendimiento de los estudiantes que presentan pruebas de conocimiento estandarizadas.

Tabla 12. Resultados de la medición del efecto

Medida <i>d</i> - Cohen	
<i>d</i> estimada	Tamaño del efecto
0,9482804	Grande

Nota: El valor *d* de la tabla se aproximó a 0,95 en el análisis por razones de espacio.

DISCUSIÓN

En la prueba Saber 11 de matemáticas, obtener entre 0 y 35 puntos ubica al estudiante en el nivel 1 de desempeño; entre 36 y 50 en el nivel 2; entre 51 y 70 en el nivel 3; y, entre 71 y 100 en el nivel 4. Basados en la medición interna de este estudio y escalando los resultados entre 0 y 100 puntos, el rendimiento en la preprueba quedó en 30,2 y en la posprueba en 44,2. Esto refleja que, aunque hubo un incremento del 14 % en el rendimiento, los estudiantes no alcanzaron, al menos, el nivel

teórico 3. Es necesario tener en cuenta que un 46 % en promedio de los estudiantes que presentaron la prueba Saber 11 entre 2017 y 2020 solo alcanzaron los niveles 1 y 2 de desempeño. Es decir, aunque el efecto del entrenamiento fue positivo, con base en la medición interna no se supera al porcentaje de estudiantes de los años de referencia mencionados.

Es importante destacar que se coincide con Díaz, et al. (2017) y Guzmán, et al. (2015) en el hecho de que los programas de entrenamiento permiten mejorar las habilidades de los alumnos para solucionar problemas de matemáticas. Aunque los resultados del presente estudio señalan que no se alcanzaron los niveles teóricos 3 o 4 presentados en el párrafo anterior, el grupo analizado mejoró en un 14 % su rendimiento después del entrenamiento.

Dado que el estudio se realizó mediante un pre experimento (muestra pareada con un solo grupo), es recomendable verificar dichos hallazgos por medio de un experimento puro de mayor complejidad, que incluya más competencias y que se contraste con los resultados oficiales de la prueba Saber 11 (validación de criterio). Los pre experimentos carecen de un adecuado grado de control de factores que pueden influir en los resultados e invalidarlos.

Como se podrá apreciar en el proceso de validación de los instrumentos en la publicación de Estrada y Guerrero (2022), estos son, en cierta medida, confiables. No obstante, es necesario complementar su proceso por validación de criterio y de constructo para su aplicación en un experimento puro. Es recomendable un mejoramiento a los ítems (preguntas) con índices de dificultad y discriminación regulares, pues sustituirlos por unos de mejor calidad puede mejorar la precisión de los resultados. Tanto los instrumentos

empleados para las pruebas, como el programa de entrenamiento, se encuentran publicados en el repositorio institucional de la UNAD, dentro del informe presentado por los autores Estrada y Guerrero (2022).

CONCLUSIONES

El ambiente virtual de aprendizaje diseñado para el programa de entrenamiento contó con un conjunto de elementos como: simulacros, materiales de apoyo y contenidos para que los estudiantes puedan explorarlos y recapitular conceptos previos. Las retroalimentaciones por cada simulacro posibilitaron la identificación de los desaciertos, los cuales podían corregir en el próximo intento.

Los estudiantes obtuvieron en promedio 7,56 puntos en la prueba previa al entrenamiento y 11,06 puntos en la prueba posterior al entrenamiento. La diferencia media es de 3,5 puntos en una escala de 0 a 25. Esto representa un incremento del 14 % en el rendimiento de los estudiantes luego de haber sido entrenados para resolver problemas y situaciones de matemáticas basados en la prueba Saber 11.

Aparentemente el programa de entrenamiento resultó ser altamente efectivo de acuerdo con la métrica de Cohen (0,95) y la categorización de la literatura propuesta. Aunque los resultados de este pre experimento son satisfactorios, es necesario entender que no se trata de un experimento puro. Los hallazgos aquí presentados se constituyen en un acercamiento al fenómeno estudiado basados en pruebas estandarizadas y procedimientos estadísticos sobre los posibles beneficios que un programa de entrenamiento puede tener en el desempeño de los estudiantes para presentar pruebas tipo Saber 11 de matemáticas.

El análisis inferencial complementario no pretende particionar el resultado conjunto obtenido ni realizar juicios de valor sobre los grupos; solo ofrece una forma adicional caracterizada para analizar los hallazgos desde otras perspectivas. En este sentido, el estudio encontró diferencias estadísticamente significativas que permiten apreciar una mayor efectividad del programa de entrenamiento para el género femenino y para estu-

diantes del establecimiento educativo de la Cruz (N.). Las causas que pudieron incidir en estas diferencias no son objeto de interpretación ni análisis en este artículo.

La edad de los estudiantes, así como el grado académico al que pertenecían (10° u 11°), son aspectos sobre los cuales el estudio no encontró diferencias significativas en la efectividad del programa de entrenamiento.

REFERENCIAS

- Arredondo, S., Diago, J., y Cañizal, A. (2010). Evaluación educativa de aprendizajes y competencias. https://eduvirtual.cuc.edu.co/moodle/pluginfile.php/647125/mod_resource/content/1/Libro%20evaluacion%20educativa%20de%20aprend%20y%20compet.pdf
- Campos, S. B., y Talavera, E. R. (2014). “El entrenamiento en estrategias sobre la comprensión lectora del enunciado del problema aritmético: un estudio empírico con estudiantes de Educación Primaria”. REOP-Revista Española de Orientación y Psicopedagogía, 17(1), 33-48. <https://doi.org/10.5944/reop.vol.17.num.1.2006.11336>
- Caughlin D. (2022). R for HR: An Introduction to Human Resource Analytics Using R. R for HR. <https://rforhr.com/>
- Caycho, T., Ventura-León, J., y Castillo-Blanco, R. (2016). “Magnitud del efecto para la diferencia de dos grupos en ciencias de la salud”. Anales del Sistema Sanitario de Navarra 39(3), 459-461. http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272016000300017&lng=es&tln_g=es
- Correa, J. C., Iral, R., y Rojas, L. (2006). “Estudio de potencia de pruebas de homogeneidad de varianza”. Revista colombiana de estadística, 29(1), 57-76. <https://www.redalyc.org/pdf/899/89929104.pdf>
- Díaz, A., Pérez, M., González, J., y Núñez, J. (2017). “Impacto de un entrenamiento en aprendizaje autorregulado en estudiantes universitarios”. Perfiles educativos, 39(157), 87-104. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982017000300087&lng=es&tln_g=es
- Estrada, J., y Guerrero, M. (2022). Diseño e implementación de un ambiente virtual de aprendizaje para el entrenamiento de estudiantes en pruebas ICFES Saber 11 de matemáticas. UNAD. <https://repository.unad.edu.co/handle/10596/51629>

- Gutiérrez, H. y Salazar, R. V. (2008). Análisis y diseño de experimentos. https://www.academia.edu/31993766/Analisis_y_dise%C3%B1o_de_experimentos_2da_edicion
- Guzmán, J., Aragón, E., Aguilar, M., Navarro, J., y Araujo, A. (2015). “Efectos de la aplicación de un programa de entrenamiento específico para el aprendizaje matemático temprano en educación infantil”. *Revista Española de Pedagogía*, 73(260) 105-119. <http://www.jstor.org/stable/24711242>
- Instituto Colombiano para la Evaluación de la Educación, ICFES (2020). Saber 11. Niveles de desempeño. Prueba de matemáticas. <https://bit.ly/3ERrm5z>
- Instituto Colombiano para la Evaluación de la Educación, ICFES (2022a). Guía de orientación saber 11 2022 – 1. <https://bit.ly/3KS1A1F>
- Instituto Colombiano para la Evaluación de la Educación, ICFES (2022b). Informe nacional de resultados del examen Saber 11° 2021. <https://bit.ly/41xJ6N8>
- López, M., y Arán, V. (2021). Transferencia de un entrenamiento de memoria de trabajo a las habilidades académicas y estrategias de resolución de problemas al inicio de la escolaridad. <http://hdl.handle.net/11336/167197>
- Mendenhall, W., Beaver, R. J. y Beaver, B. M (2010). Introducción a la probabilidad y estadística. <https://www.fcfm.buap.mx/jzacarias/cursos/estad2/libros/book5e2.pdf>
- Meza, C. (2012). “¿Entrenar los cuerpos es educar los sujetos? Un acercamiento a la construcción pedagógica de sujetos/cuerpos en la Licenciatura en Arte Teatral”. *Papel Escena*, (11), 110-123. <https://revistas.bellasartes.edu.co/index.php/papel/article/download/180/150>
- Quintana, L., y Mendoza, M. A. (2016). Econometría aplicada utilizando R. http://saree.com.mx/econometriaR/sites/default/files/Ebook_econometriaR.pdf
- Sampieri, R., Fernández, C., y Lucio, P. (2014). Metodología de la investigación. 6ta edición. <https://www.uca.ac.cr/wp-content/uploads/2017/10/Investigacion.pdf>
- Schunk, D. (2010). Teorías del aprendizaje. <https://fundasira.cl/wp-content/uploads/2017/03/TEORIAS-DEL-APRENDIZAJE.-DALE-SCHUNK..pdf>
- Tapia, C. E. F., y Cevallos, K. L. F. (2021). “Pruebas para comprobar la normalidad de datos en procesos productivos: Anderson-Darling, Ryan-Joiner, Shapiro-Wilk y Kolmogórov-Smirnov”. *Societas*, 23(2), 83-106. <https://revistas.up.ac.pa/index.php/societas/article/view/2302>
- Walpole, R. E., Myers, R. H., y Myers, S. L. (1999). Probabilidad y estadística para ingenieros. <https://bit.ly/41CmLOX>