Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), enero-febrero 2026,
Volumen 10, Número 1.

https://doi.org/10.37811/cl_rcm.v10i1

EVALUACIÓN DE LA PLANIFICACIÓN DIDÁCTICA POR INTELIGENCIA ARTIFICIAL GENERATIVA Y SU IMPACTO EN EL RENDIMIENTO ACADÉMICO EN LA NUEVA ESCUELA MEXICANA CON ENFOQUE EN NECESIDADES EDUCATIVAS ESPECIALES: ESTUDIO DE CASO EN SINALOA, MÉXICO PARA LA ASIGNATURA DE EDUCACIÓN FÍSICA

EVALUATION OF DIDACTIC PLANNING BY GENERATIVE ARTIFICIAL INTELLIGENCE AND ITS IMPACT ON ACADEMIC PERFORMANCE IN THE NEW MEXICAN SCHOOL WITH A FOCUS ON SPECIAL EDUCATIONAL NEEDS: A CASE STUDY IN SINALOA, MEXICO FOR THE SUBJECT OF PHYSICAL EDUCATION

Mónica de Jesús López Morales
Universidad Autónoma de Sinaloa

Mónica del Carmen Morales Parra
Universidad Autónoma de Sinaloa

Ernesto López Morales
Universidad Autónoma de Sinaloa

Sarahi Herrera Hernández
Universidad Autónoma de Sinaloa

Pedro Antonio Valdez Lizárraga
Universidad Autónoma de Sinaloa

DOI: https://doi.org/10.37811/cl_rcm.v10i1.22727

Evaluación de la planificación didáctica por inteligencia artificial generativa y su impacto en el rendimiento académico en la Nueva Escuela Mexicana con enfoque en Necesidades Educativas Especiales: Estudio de caso en Sinaloa, México para la asignatura de Educación Física

Dra. Mónica de Jesús López Morales[1]
moniicastarr@gmail.com
https://orcid.org/0000-0001-9774-3676
Universidad Autónoma de Sinaloa

Dra. Mónica del Carmen Morales Parra
netomonii@gmail.com
https://orcid.org/0009-0005-3038-1398
Universidad Autónoma de Sinaloa

Dr. Ernesto López Morales
netoman117@gmail.com
https://orcid.org/0009-0008-0979-3170
Universidad Autónoma de Sinaloa

Dra. Sarahi Herrera Hernández
sarahi.herrera@uas.edu.mx
https://orcid.org/0009-0007-5501-7175
Universidad Autónoma de Sinaloa

M.C. Pedro Antonio Valdez Lizárraga
Pedro.valdez@uas.edu.mx
https://orcid.org/0000-0001-5773-4628
Universidad Autónoma de Sinaloa

RESUMEN

La presente investigación analiza la relación entre la calidad de la planificación didáctica, el uso de Inteligencia Artificial Generativa (IAG) y el rendimiento académico en Educación Física dentro del contexto de la Nueva Escuela Mexicana (NEM). Se adoptó un enfoque cuantitativo con diseño correlacional-explicativo, aplicando cuestionarios y rúbricas validados ($\alpha = 0.91$) a una muestra estratificada de 323 docentes de Sinaloa, México (Mazatlán, Culiacán y Los Mochis). Los resultados revelaron una alta asociación entre la calidad de la planificación y las calificaciones promedio ($r = 0.994$, $p < 0.001$), estableciéndose como el principal predictor del desempeño. Aunque el uso directo de herramientas como GPT no mostró un impacto significativo en las calificaciones ($p = 0.723$), sí evidenció una fuerte correlación con la adaptación docente al modelo NEM ($r_s = 0.971$). Se concluye que, si bien la IAG no sustituye la estructura pedagógica, actúa como un facilitador clave para la innovación y la alineación con las políticas educativas actuales, optimizando procesos sin ser la causa directa de la mejora sumativa en el rendimiento estudiantil.

Palabras clave: planificación didáctica; inteligencia artificial generativa; Nueva Escuela Mexicana; rendimiento académico; educación física

[1] Autor principal
Correspondencia: moniicastarr@gmail.com

Evaluation of didactic planning by generative artificial intelligence and its impact on academic performance in the New Mexican School with a focus on Special Educational Needs: A case study in Sinaloa, Mexico for the subject of Physical Education

ABSTRACT

This research analyzes the relationship between the quality of didactic planning, the use of Generative Artificial Intelligence (GAI), and academic performance in Physical Education within the context of the New Mexican School (NEM). A quantitative approach with a correlational–explanatory design was adopted, applying validated questionnaires and rubrics ($\alpha = .91$) to a stratified sample of 323 teachers from Sinaloa, Mexico (Mazatlán, Culiacán, and Los Mochis). The results revealed a very high positive association between planning quality and students' average grades ($r = .994$, $p < .001$), establishing it as the main predictor of performance in this sample. Although the direct use of tools such as GPT did not show a significant impact on grades ($p = .723$), it did show a strong correlation with teachers' adaptation to the NEM model ($r_s = .971$). It is concluded that, although GAI does not replace pedagogical structure, it acts as a key facilitator for innovation and alignment with current educational policies, optimizing processes without being the direct cause of summative improvement in student performance.

Keywords: didactic planning; generative artificial intelligence; New Mexican School; academic performance; physical education

Artículo recibido 10 diciembre 2025
Aceptado para publicación: 10 enero 2026

INTRODUCCIÓN

La incorporación de la Inteligencia Artificial Generativa (IAG) en el espectro educativo contemporáneo constituye uno de los retos epistemológicos y pedagógicos de mayor trascendencia en la última década, superando en complejidad a disrupciones tecnológicas previas como la digitalización de contenidos o los sistemas de gestión del aprendizaje (LMS). Investigaciones recientes, como las de García-Peñalvo et al. (2024) y Gallent-Torres et al. (2023), postulan que el advenimiento de modelos de lenguaje masivo (LLM), ejemplificados por herramientas como GPT, no solo reconfigura los mecanismos de producción y consumo de contenidos académicos, sino que interpela las dinámicas tradicionales inherentes al proceso de enseñanza-aprendizaje.

Esta tecnología desplaza al docente de su rol histórico como único poseedor del conocimiento, obligándolo a transitar hacia una función de mediador crítico y diseñador de experiencias de aprendizaje situado. No obstante, si bien la literatura internacional abunda en análisis teóricos concernientes a las implicaciones éticas, la integridad académica y la personalización del aprendizaje en la educación superior (Ojeda et al., 2023), se advierte una carencia sustancial de evidencia empírica que cuantifique el impacto tangible de estas herramientas en el nivel de educación básica. Esta laguna es particularmente notoria en disciplinas de naturaleza práctica y vivencial como la Educación Física, donde la mediación tecnológica parece, a primera vista, distante de la praxis corporal y el desarrollo psicomotor.

El panorama educativo mexicano experimenta actualmente una metamorfosis paradigmática derivada de la instrumentación de la Nueva Escuela Mexicana (NEM), un modelo que jerarquiza la equidad, la inclusión, la interculturalidad crítica y el desarrollo integral del discente bajo un enfoque eminentemente humanista (Secretaría de Educación Pública [SEP], 2022). Dicho modelo exige del cuerpo docente una competencia de planificación didáctica altamente adaptativa, facultada para atender la heterogeneidad cognitiva, motriz y sociocultural del alumnado, identificando y minimizando las Barreras para el Aprendizaje y la Participación (BAP) presentes en los contextos escolares.

En esta coyuntura, la planificación didáctica trasciende su carácter meramente administrativo o burocrático para consolidarse como el eje rector del desempeño académico y la concreción curricular. La hipótesis subyacente en la praxis educativa contemporánea sugiere que la tecnología, específicamente la IAG, podría fungir como un catalizador instrumental para dicha adaptación,

automatizando tareas rutinarias de diseño instruccional y permitiendo al docente focalizar su atención en la interacción pedagógica sustantiva. Sin embargo, prevalece la incertidumbre respecto a si esta eficiencia tecnológica en la fase de diseño se traduce efectivamente en una mejora sumativa del rendimiento estudiantil o si, por el contrario, su incidencia se circunscribe exclusivamente a la optimización de los tiempos de gestión docente.

La Educación Física, disciplina frecuentemente relegada en los estudios de tecnología educativa los cuales suelen privilegiar áreas STEM (Ciencia, Tecnología, Ingeniería y Matemáticas), confronta desafíos singulares en este nuevo ecosistema digital. La naturaleza kinestésica de la asignatura, aunada a la necesidad imperiosa de personalizar el desarrollo de habilidades motrices y competencias socioemocionales en entornos diversos, demanda una estructuración pedagógica rigurosa y creativa, donde la IAG podría facilitar el diseño de variantes motrices y ajustes razonables para estudiantes con diversas capacidades, una tarea que tradicionalmente consume gran parte del tiempo lectivo docente. Evidencia previa (Carriazo Díaz et al., 2020) ha corroborado que una planificación deficiente constituye un predictor significativo de bajo rendimiento y desmotivación escolar, independientemente de los recursos tecnológicos disponibles.

Este fenómeno adquiere matices específicos en el estado de Sinaloa, una entidad caracterizada por una marcada heterogeneidad estructural entre sus principales núcleos urbanos (Mazatlán, Culiacán y Los Mochis) y sus zonas periféricas. Esta disparidad convierte a la región en un laboratorio idóneo para analizar si la tecnología actúa como un ecualizador de oportunidades docentes o como un amplificador de desigualdades preexistentes. La problemática de investigación, por tanto, no se circunscribe a la capacidad técnica inherente a la IA, sino que converge en su interacción dialéctica con la praxis pedagógica situada. Se observa en el discurso educativo actual una tendencia hacia un "tecno-optimismo" acrítico que presupone que la mera incorporación de herramientas digitales eleva *per se* la calidad educativa (Silva Hernández & Martínez Prats, 2021). No obstante, la observación empírica sugiere una realidad más matizada: la tecnología desprovista de una estructura pedagógica sólida y una intencionalidad didáctica clara resulta inocua.

En consecuencia, resulta imperativo dilucidar empíricamente si el empleo de asistentes virtuales como GPT influye directamente en las calificaciones de los estudiantes como indicador de rendimiento

sumativo o si su valor primario reside en facilitar la adaptación del docente a los exigentes principios de flexibilidad e inclusión de la NEM. En virtud de lo anterior, el objetivo general de esta investigación fue analizar la relación entre la calidad de la planificación didáctica, la percepción y uso de herramientas de IAG, y su impacto en el rendimiento académico y la adaptación al modelo de la Nueva Escuela Mexicana en docentes de educación física de la región. De manera específica, se buscó determinar la existencia de diferencias significativas en el desempeño estudiantil entre docentes usuarios y no usuarios de GPT, así como evaluar en qué medida esta herramienta tecnológica contribuye a la materialización operativa de los principios de inclusión y equidad exigidos por el sistema educativo nacional.

METODOLOGÍA

Enfoque y Diseño de Investigación

La presente investigación se fundamenta epistemológicamente en el paradigma positivista, adoptando un enfoque cuantitativo de corte racionalista. Esta postura asume la premisa de que la realidad educativa, si bien compleja, posee dimensiones objetivas y observables que son susceptibles de medición precisa y cuantificación rigurosa. En consecuencia, se privilegió la recolección sistemática de datos numéricos y su posterior tratamiento mediante técnicas estadísticas avanzadas, permitiendo no solo la descripción de fenómenos, sino la contrastación empírica de hipótesis para establecer patrones de regularidad que trasciendan la subjetividad individual del docente.

En cuanto al alcance, el estudio se definió estratégicamente como correlacional-explicativo, una dualidad metodológica que permitió abordar el fenómeno en dos niveles de profundidad complementarios. En una primera instancia, la dimensión correlacional se orientó a cuantificar la magnitud, dirección y significancia estadística de la asociación entre variables críticas: la calidad técnica de la planificación didáctica, la frecuencia y tipo de uso de la Inteligencia Artificial Generativa (IAG), y los indicadores de rendimiento académico. Posteriormente, el nivel explicativo buscó trascender la mera descripción asociativa para desentrañar los mecanismos de influencia causal subyacentes, identificando específicamente a la IAG no como un fin en sí mismo, sino como una variable interviniente crucial que modula y facilita la adaptación docente a los exigentes lineamientos normativos y pedagógicos de la Nueva Escuela Mexicana (NEM).

El diseño implementado fue no experimental, transversal y retrospectivo. Al renunciar a la manipulación deliberada de variables independientes (típica de los diseños experimentales), se privilegió el análisis de los fenómenos en su contexto ecológico natural, observando las prácticas docentes y los resultados estudiantiles tal como ocurren en la realidad escolar cotidiana, sin las distorsiones artificiales de un entorno de laboratorio. La recolección de datos en un corte temporal único permitió capturar una "fotografía" diagnóstica del estado actual de la integración tecnológica y la praxis pedagógica, asegurando así la validez externa de los hallazgos y su potencial generalización a contextos educativos con características sociodemográficas similares (Hernández-Sampieri & Mendoza, 2018).

Población y Muestra

La población objetivo comprendió la totalidad de los docentes de educación básica (niveles de primaria y secundaria) adscritos al sistema educativo del estado de Sinaloa, México. Para garantizar la representatividad estadística y el control de variables intervinientes, se utilizó un muestreo probabilístico estratificado que derivó en una muestra final de 323 docentes (N=323). La estratificación se ejecutó bajo dos criterios rectores: la ubicación geográfica, asegurando una proporción acorde a la densidad educativa de las regiones neurálgicas del estado (Mazatlán, Culiacán y Los Mochis), y el perfil tecnológico, estableciendo una asignación equilibrada *a priori* entre docentes usuarios de GPT y no usuarios. Esta segmentación, detallada en la Tabla 1, fue fundamental para mitigar sesgos de selección y permitir comparaciones robustas entre grupos.

Tabla 1. Distribución Estratificada de la Muestra Docente

Ubicación Geográfica	Frecuencia (n)	Porcentaje (%)	Usuarios GPT (n)	No Usuarios GPT (n)
Mazatlán	194	60.1%	97	97
Culiacán	65	20.1%	33	32
Los Mochis	64	19.8%	32	32
Total	**323**	**100%**	**162**	**161**

Nota. Datos derivados de la matriz de análisis del estudio.

Como se aprecia en la Tabla 1, la distribución geográfica de la muestra refleja la concentración demográfica y educativa del estado, con una participación predominante de la región de Mazatlán

(n=194), que constituye el 60.1% del total. Las regiones de Culiacán y Los Mochis presentan una representación equiparable, con el 20.1% y 19.8% respectivamente. Es fundamental destacar la rigurosidad en la asignación del perfil tecnológico, logrando un equilibrio casi simétrico entre los docentes usuarios de GPT (n=162) y los no usuarios (n=161) a nivel global. Esta paridad se mantiene consistente al interior de cada estrato geográfico, lo que garantiza que las comparaciones estadísticas entre ambos grupos no se vean sesgadas por factores de ubicación, validando así la homogeneidad estructural de los subgrupos de análisis.

Técnicas e Instrumentos de Recolección

La estrategia de recolección de datos se instrumentó mediante el diseño y aplicación de una batería de tres herramientas complementarias, sometidas a un riguroso proceso de validación de contenido por juicio de expertos y análisis de fiabilidad. El análisis psicométrico global arrojó un coeficiente de consistencia interna Alfa de Cronbach de 0.91, evidenciando una homogeneidad excelente en la medición de los constructos estudiados.

En primera instancia, se aplicó el *Cuestionario Docente sobre Desempeño Estudiantil*, un instrumento *ad hoc* estructurado para cuantificar la percepción experta del profesorado. Este cuestionario emplea una escala Likert de cinco puntos (donde 1 representa "Deficiente" y 5 "Sobresaliente") para evaluar no solo el rendimiento académico sumativo, sino también indicadores cualitativos críticos como la motivación, la autonomía y, fundamentalmente, la alineación de las competencias desarrolladas con los ejes articuladores de la Nueva Escuela Mexicana (NEM).

De manera complementaria, y con el objetivo de objetivar la evaluación del rendimiento en una disciplina práctica como la Educación Física, se implementó la *Rúbrica Estandarizada de Desempeño Escolar*. Este instrumento permite trascender la asignación de una calificación numérica aislada, desglosando el desempeño en cinco dimensiones holísticas: dominio del contenido (saberes teóricos), habilidades prácticas (ejecución motriz), participación, organización y desarrollo socioemocional. Esta desagregación resulta vital para identificar en qué áreas específicas incide la planificación didáctica.

Finalmente, para robustecer la validez interna mediante la triangulación de fuentes y controlar el sesgo de deseabilidad social inherente a los autorreportes, se empleó la *Guía de Observación Estandarizada*. Este instrumento se utilizó para registrar comportamientos *in situ* durante las sesiones prácticas. La

fiabilidad de los registros fue garantizada mediante el cálculo del Coeficiente kappa de Cohen (κ) para la concordancia inter-evaluador, obteniendo un valor superior a 0.85 ($\kappa > 0.85$). La Tabla 2 presenta la operacionalización técnica de las variables analizadas.

Tabla 2. Matriz de Operacionalización de Variables y Técnicas de Recolección

Variable	Definición Operacional	Instrumento	Nivel de Medición
Calidad de Planificación	Grado de alineación con objetivos de aprendizaje y estructura pedagógica.	Rúbrica de Evaluación	Escala de Razón (Puntaje)
Uso de GPT	Frecuencia e integración de IAG en el diseño instruccional.	Cuestionario Docente	Nominal Dicotómica (Sí/No)
Rendimiento Académico	Promedio de calificaciones y nivel de competencia motriz alcanzado.	Registros Escolares / Rúbrica	Escala de Razón / Ordinal
Adaptación a NEM	Nivel de incorporación de principios de inclusión y equidad.	Cuestionario Docente	Escala Ordinal (Likert)
Participación Estudiantil	Frecuencia de interacción activa en sesiones prácticas.	Guía de Observación	Escala Ordinal

Nota. Elaboración propia basada en el diccionario de variables del estudio.

La Tabla 2 detalla la estrategia de instrumentalización que permitió transformar constructos teóricos complejos en datos empíricos procesables. Es imperativo señalar la naturaleza multidimensional de la variable dependiente "Rendimiento Académico", la cual no se limitó a la calificación sumativa tradicional (escala de razón), sino que integró una valoración cualitativa ordinal a través de la rúbrica, permitiendo capturar matices del desarrollo competencial que una simple nota numérica podría enmascarar. Por su parte, la variable independiente "Calidad de Planificación" fue tratada como una variable continua de razón, derivada de la sumatoria de puntajes de la rúbrica de evaluación, lo que habilitó la aplicación de pruebas paramétricas robustas como la correlación de Pearson. Finalmente, la operacionalización dicotómica del "Uso de GPT" fue una decisión metodológica clave para establecer grupos de comparación nítidos, facilitando el análisis de diferencias de medias (prueba *t*) sin la ambigüedad de grados intermedios de uso.

Procedimiento

El despliegue operativo de la investigación se articuló rigurosamente en tres fases secuenciales. Inicialmente, se ejecutó una fase de validación de contenido y constructo mediante un estudio piloto con un grupo control de 30 docentes (n=30) ajenos a la muestra final. Esta etapa fue crucial no solo para calibrar la redacción de los ítems y eliminar posibles ambigüedades semánticas, sino también para confirmar preliminarmente las propiedades psicométricas de los instrumentos, asegurando que la terminología utilizada fuera pertinente al contexto educativo sinaloense. Posteriormente, el levantamiento de datos se desplegó a través de una estrategia diseñada para maximizar la cobertura territorial.

Por un lado, se utilizaron plataformas digitales (Google Forms) para agilizar la recolección en docentes de zonas urbanas con acceso estable a internet. Esta metodología dual fue fundamental para garantizar la inclusión de docentes de diversos estratos socioeconómicos, mitigando efectivamente el sesgo de la brecha digital y asegurando la representatividad de las tres regiones estudiadas. Finalmente, la información recabada fue sometida a un riguroso proceso de sistematización. Se realizaron procedimientos de depuración para identificar y corregir valores atípicos o inconsistencias, así como un control de calidad para descartar cuestionarios incompletos. Una vez validada la integridad de los datos, estos fueron codificados e integrados en una matriz de análisis general, preparándolos para su procesamiento estadístico.

Análisis de Datos

El tratamiento estadístico de la información se llevó a cabo utilizando el software especializado IBM SPSS Statistics versión 27, siguiendo un protocolo analítico jerarquizado diseñado para maximizar la validez interna de los hallazgos. En la fase inicial, se ejecutó un análisis descriptivo exhaustivo con el propósito de caracterizar el comportamiento distributivo de las variables clave; para ello, se calcularon medidas de tendencia central (media, mediana) y de dispersión (desviación estándar), fundamentales para establecer una línea base de la muestra y detectar la presencia de valores atípicos (*outliers*) que pudieran sesgar los resultados posteriores. Para la contrastación de hipótesis comparativas, se seleccionó la prueba paramétrica t de Student para muestras independientes, previa verificación rigurosa de los supuestos de normalidad (prueba de Kolmogorov-Smirnov) y homogeneidad de varianzas (prueba de

Levene). Esta técnica permitió determinar con precisión la existencia o ausencia de diferencias estadísticamente significativas en las calificaciones promedio entre el grupo de docentes usuarios de GPT y el de no usuarios, aislando el efecto de la variable tecnológica.

La estructura asociativa de los datos se exploró mediante un enfoque correlacional dual: se utilizó el coeficiente de correlación de Pearson (r) para examinar la linealidad en datos continuos de razón (como la relación entre puntajes de planificación y calificaciones), y el coeficiente de Spearman (r_s) para variables de naturaleza ordinal, tales como las escalas de percepción tipo Likert. Posteriormente, con el objetivo de determinar el peso predictivo específico de cada variable independiente sobre el rendimiento académico, se construyeron modelos de regresión lineal múltiple. La robustez de estos modelos fue validada diagnosticando la multicolinealidad a través del cálculo del Factor de Inflación de la Varianza (VIF), asegurando que las estimaciones de los coeficientes no estuvieran infladas por correlaciones excesivas entre predictores. Finalmente, para evaluar la significancia de los cambios intra-sujeto en el rendimiento pre y post intervención (implementación de planificación estructurada), se empleó la prueba de rangos de Wilcoxon, seleccionada por su potencia estadística en el análisis de muestras relacionadas no paramétricas, ideal para detectar cambios sutiles pero consistentes en el desempeño longitudinal.

Consideraciones Éticas y Limitaciones

El estudio se apegó estrictamente a los lineamientos éticos internacionales estipulados en la Declaración de Helsinki. Se implementó un protocolo riguroso de consentimiento informado, garantizando a cada participante el pleno conocimiento de los objetivos del estudio, la naturaleza voluntaria de su colaboración y el derecho inalienable a la revocación de su participación en cualquier etapa del proceso. Para asegurar la integridad de los datos, se establecieron mecanismos de anonimizarían mediante códigos alfanuméricos, protegiendo la identidad de los docentes y sus centros escolares.

En el ámbito de las limitaciones metodológicas, el análisis de regresión evidenció un Factor de Inflación de la Varianza (VIF) de 12.435 entre la calidad de la planificación y el rendimiento académico. Si bien este valor excede los umbrales convencionales de no colinealidad, no se interpreta únicamente como un error técnico del modelo, sino también como un hallazgo estructural que refleja una interdependencia conceptual casi absoluta entre la fase de diseño didáctico y la ejecución pedagógica en el marco de la NEM, lo que dificulta aislar el efecto "puro" de cada variable. Adicionalmente, es imperativo

circunscribir la generalización de los hallazgos al contexto socioeducativo del estado de Sinaloa. Las particularidades regionales, culturales y de infraestructura tecnológica de la muestra sinaloense constituyen un estudio de caso específico que, si bien ofrece una alta validez ecológica interna, requiere cautela al extrapolarse a otras entidades federativas con dinámicas educativas disímiles.

RESULTADOS

El análisis de los datos se estructuró estratégicamente mediante una arquitectura analítica secuencial y progresiva, diseñada no solo para responder con precisión matemática a las preguntas de investigación planteadas, sino para capturar la complejidad fenomenológica de la práctica docente actual. Este diseño metodológico evalúa la interacción sistémica y multifactorial entre tres ejes críticos: la calidad técnica y pedagógica de la planificación didáctica, la mediación tecnológica facilitada por la adopción de herramientas de Inteligencia Artificial Generativa (IAG) y, como variable dependiente final, el desempeño académico en la asignatura de Educación Física. Esta aproximación trasciende la simple presentación de cifras aisladas o tablas descontextualizadas; busca, en cambio, construir una narrativa estadística coherente que desentrañe los mecanismos subyacentes del rendimiento escolar. En el contexto de la Nueva Escuela Mexicana (NEM), donde se priorizan procesos formativos sobre resultados estandarizados, este análisis busca identificar si la tecnología actúa como un facilitador instrumental de la gestión docente o si posee la capacidad de transformar sustantivamente los resultados de aprendizaje.

A continuación, se presentan los hallazgos estadísticos organizados jerárquicamente por dimensiones de análisis, integrando en un mismo marco interpretativo la potencia de la estadística descriptiva y la profundidad de la inferencial. En una primera instancia, el análisis descriptivo permite caracterizar el "estado del arte" de las variables en las aulas sinaloenses, estableciendo una línea base sobre cómo planifican los docentes y cómo rinden los estudiantes. Posteriormente, la estadística inferencial (pruebas t, correlaciones y regresiones) se utiliza para contrastar hipótesis y establecer relaciones de causalidad, permitiendo discernir entre correlaciones espurias y vínculos pedagógicos reales. Este enfoque dual es fundamental para no solo describir qué está ocurriendo (por ejemplo, los promedios de calificación), sino para explicar por qué y cómo interactúan los factores pedagógicos y tecnológicos. De esta manera, se proporciona una base empírica sólida para la discusión posterior, permitiendo distinguir entre el

impacto real de la planificación estructurada y el "ruido" o falsas expectativas generadas por la mera incorporación de herramientas digitales como GPT.

3.1. Caracterización Descriptiva de las Variables

En una fase preliminar y fundamental del análisis, se procedió a examinar el comportamiento distribucional de las variables centrales del estudio. Este paso no solo tuvo como objetivo establecer una línea base estadística, sino también diagnosticar la naturaleza de la muestra docente y estudiantil en el contexto de la reforma educativa actual. La Tabla 3 presenta un compendio detallado de los estadísticos descriptivos para la totalidad de la muestra (N = 323), desglosando medidas de tendencia central, dispersión y forma que revelan patrones pedagógicos subyacentes.

Al analizar el Promedio de Calificaciones, se observa una media aritmética de 8.12 en una escala estándar de 0 a 10. Este valor, superior al punto medio teórico, indica un nivel de rendimiento general satisfactorio. Sin embargo, la desviación estándar ($\sigma = 1.32$) sugiere una heterogeneidad moderada en los resultados, implicando que, si bien el promedio es alto, existen brechas de desempeño notables dentro de los grupos escolares. La asimetría negativa (-0.14) corrobora esta tendencia: la distribución de las notas no es perfectamente simétrica, sino que se inclina ligeramente hacia el extremo superior, lo que es característico de sistemas de evaluación en educación básica donde las tasas de reprobación suelen ser bajas.

Por otro lado, la variable Calidad de la Planificación arroja resultados particularmente reveladores. Con una media de 92.40 sobre 100 y una asimetría marcadamente negativa (-0.89), nos encontramos ante un evidente "efecto techo". Esto significa que la gran mayoría de los docentes, independientemente de su perfil tecnológico o ubicación geográfica, reporta un cumplimiento casi total de los criterios normativos de planificación. Este hallazgo sugiere que el magisterio sinaloense posee una alta adherencia formal a los lineamientos institucionales, lo cual plantea un escenario interesante: si la planificación es uniformemente excelente, las variaciones en el rendimiento estudiantil deben explicarse por factores ajenos a la estructura del diseño instruccional, o bien, la rúbrica de evaluación de la planificación podría estar capturando cumplimiento burocrático más que innovación pedagógica.

Adicionalmente, la variable Tiempo de Planificación muestra una media de 115.3 minutos semanales con una desviación estándar considerable ($\sigma = 28.6$). Esta dispersión indica que, aunque la calidad del

producto final (el plan didáctico) es homogénea, el "costo temporal" para lograrlo varía drásticamente entre docentes, lo que abre la puerta a la hipótesis de que herramientas como GPT podrían estar optimizando la eficiencia (reduciendo tiempo) sin necesariamente alterar la calidad formal del documento. Finalmente, la Adaptación al Modelo NEM (4.15/5) y la Percepción de Utilidad de GPT (4.38/5) reflejan una disposición actitudinal muy positiva hacia la innovación y la reforma educativa.

Tabla 3. Estadísticos Descriptivos de las Variables de Estudio

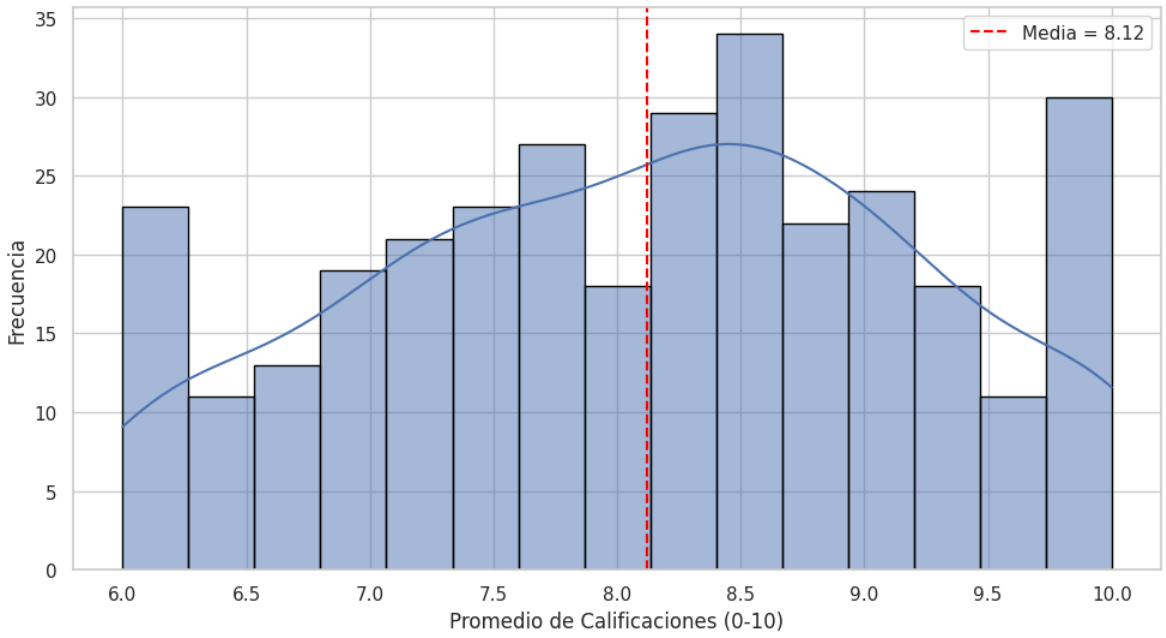
Variable	N	Mínimo	Máximo	Media ($\bar{x}$)	Desv. Est. (σ)	Asimetría
Promedio de Calificaciones (0-10)	323	6.50	9.80	8.12	1.32	-0.14
Calidad de Planificación (Puntaje)	323	65.0	100.0	92.40	8.15	-0.89
Tiempo de Planificación (min/sem)	323	45.0	180.0	115.3	28.6	0.45
Adaptación al Modelo NEM (1-5)	323	2.0	5.0	4.15	0.82	-0.67
Percepción de Utilidad GPT (1-5)	162	1.0	5.0	4.38	0.76	-1.02

Nota. La variable "Percepción de Utilidad GPT" se calculó exclusivamente sobre el subgrupo de usuarios (n = 162). Datos procesados en SPSS v27.

Los datos descriptivos revelan un escenario pedagógico caracterizado por un desempeño docente alto en términos de planificación formal (Media = 92.40), lo que contrasta con una mayor variabilidad en el rendimiento estudiantil (σ = 1.32). La asimetría negativa en ambas variables indica una tendencia hacia puntuaciones altas, sugiriendo que la mayoría de la muestra se sitúa en niveles de competencia satisfactorios. Es relevante notar que la percepción de utilidad del GPT es muy positiva (4.38/5), lo que anticipa una buena recepción de la tecnología, aunque esto no necesariamente se traduzca en una mejora directa de las calificaciones, como se verá en los análisis inferenciales.

Para visualizar la distribución del rendimiento académico, la Figura 1 ilustra la frecuencia de las calificaciones. Se evidencia una tendencia leptocúrtica con un sesgo negativo leve, indicando una concentración de estudiantes con calificaciones superiores a la media teórica (8.0), lo que es consistente con las dinámicas de evaluación en Educación Física donde la participación activa pondera significativamente en la nota final.

Figura 1. Histograma de Distribución de Calificaciones



Nota. elaboración Propia.

El histograma confirma visualmente la tendencia hacia el alto rendimiento observada en los descriptivos. La concentración de datos en el rango de 7.5 a 9.0, con una cola izquierda menos pronunciada, sugiere que el sistema de evaluación en Educación Física tiende a premiar el cumplimiento y la participación, generando un "techo" en las calificaciones que podría enmascarar diferencias sutiles en el aprendizaje competencial. Esta distribución no normal justifica el uso de pruebas robustas y análisis complementarios para detectar efectos reales de las variables independientes.

Análisis Comparativo: El Impacto Diferencial del Uso de GPT

Uno de los objetivos nodales y más críticos que guiaron el diseño de esta investigación consistió en dilucidar, con base en evidencia empírica y más allá de las percepciones subjetivas, si la incorporación operativa de herramientas de Inteligencia Artificial Generativa (IAG) como GPT ejerce una influencia causal directa, cuantificable y significativa sobre los resultados sumativos de aprendizaje de los estudiantes. Para abordar esta interrogante con el debido rigor metodológico, se procedió a la aplicación de la prueba paramétrica t de Student para muestras independientes. Este procedimiento estadístico fue seleccionado por su robustez para comparar las medias de dos grupos poblacionales distintos en este caso, la cohorte de docentes usuarios de la tecnología frente a la cohorte de no usuarios asumiendo la independencia de las observaciones y la homogeneidad de las varianzas.

El análisis detallado de los resultados, los cuales se presentan desglosados en la Tabla 4, arroja una evidencia estadística contundente que desafía ciertas expectativas intuitivas: no existen diferencias estadísticamente significativas en el rendimiento académico directo entre ambos grupos (t (321) = -0.354, p = 0.723). El valor p obtenido, al situarse considerablemente por encima del umbral de significancia estándar ($\alpha = 0.05$), obliga a aceptar la hipótesis nula de igualdad de medias. En términos concretos, la diferencia observada en los promedios es marginal y pedagógicamente irrelevante (apenas 0.04 puntos en una escala decimal), lo que constituye un hallazgo de gran trascendencia para la discusión educativa actual.

Este dato permite rechazar con firmeza la hipótesis tecnocéntrica o "solucionista" que postula que la mera integración de tecnología avanzada per se eleva automáticamente las calificaciones. Por el contrario, los datos sugieren que la IAG actúa como una herramienta de soporte al proceso de enseñanza, pero no como un determinante directo del producto evaluativo final, reafirmando que la calidad del aprendizaje sigue estando mediada por factores didácticos más profundos que la simple eficiencia instrumental.

Tabla 4. Prueba t de Student para Diferencias de Medias en Rendimiento Académico según Uso de GPT

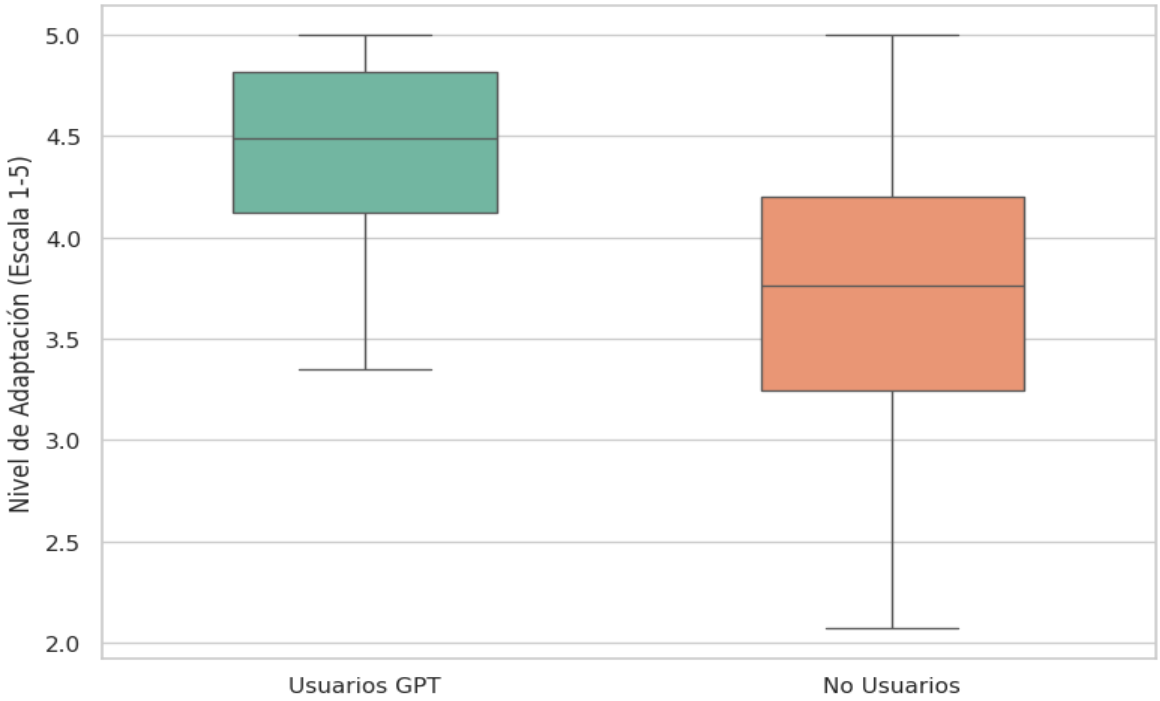
Grupo	n	Media ($\bar{x}$)	Desv. Est.	t	gl	Sig. (p)	Diferencia de medias	Error típ. de la diferencia
Usuarios GPT	162	8.14	1.28	-0.354	321	0.723	0.04	0.113
No Usuarios	161	8.10	1.35					

Nota. Nivel de significancia establecido en $\alpha = 0.05$. El valor $p > 0.05$ indica que se acepta la hipótesis nula de igualdad de medias.

La ausencia de significancia estadística ($p = 0.723$) es un hallazgo contundente que desafía la narrativa tecno-optimista. Indica que la mera adopción de herramientas de IA por parte del docente no genera un impacto inmediato y directo en la calificación final del estudiante. Esto sugiere que el valor de la tecnología no reside en "inflar" notas, sino posiblemente en optimizar procesos intermedios (como la planificación o la adaptación curricular) que no siempre se reflejan en el promedio sumativo a corto plazo.

Sin embargo, el análisis de variables cualitativas arroja una perspectiva diferente. Al comparar la Adaptación al Modelo NEM mediante la prueba U de Mann-Whitney, se encontró una diferencia altamente significativa ($Z = -4.12$, $p < 0.001$). Como se aprecia visualmente en la Figura 2, los docentes usuarios de GPT reportan medianas superiores en la escala de adaptación (4.5 vs 3.8), sugiriendo que la herramienta actúa como un andamiaje eficaz para interpretar e implementar los complejos principios pedagógicos de la reforma educativa.

Figura 2. Diagrama de Cajas Comparativo de Adaptación a la NEM



Nota. elaboración Propia.

El diagrama de cajas revela una diferencia clara y sistemática en la percepción de adaptación al nuevo modelo educativo. Mientras que el grupo de "No Usuarios" muestra una mayor dispersión y medianas más bajas (indicando dificultades o incertidumbre en la implementación de la NEM), el grupo de "Usuarios GPT" presenta una distribución más compacta y elevada hacia los valores superiores de la escala. Esto visualiza el rol de la IA como un "facilitador cognitivo" para el docente, ayudándole a traducir los lineamientos teóricos de la reforma en planificaciones prácticas y alineadas.

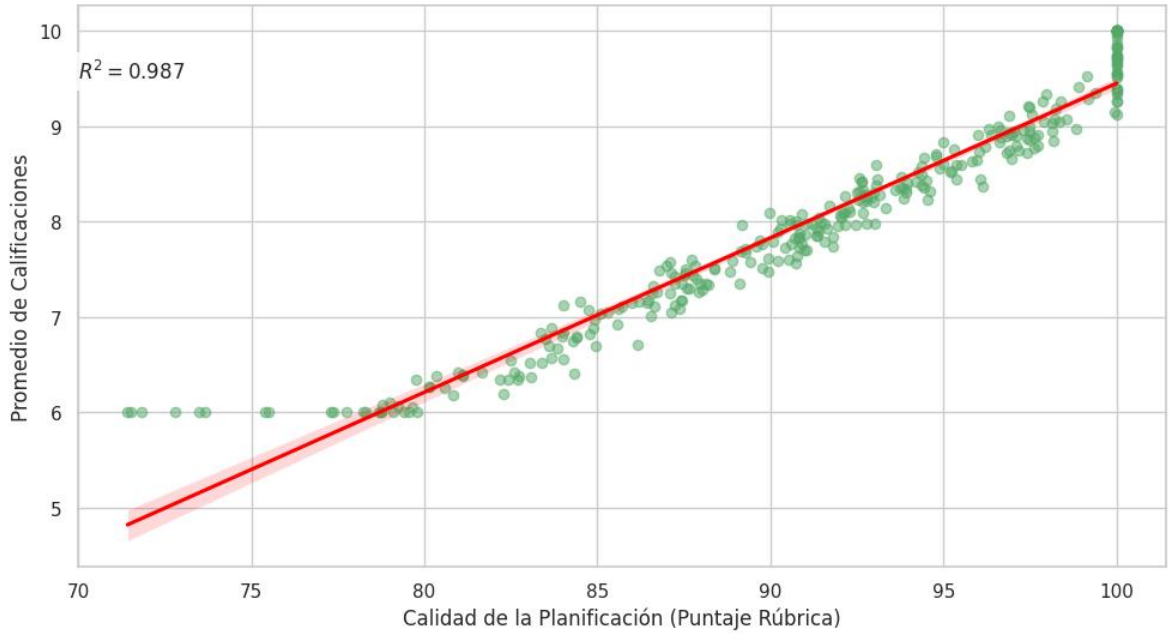
Análisis Correlacional y Modelos Explicativos

Tras la caracterización individual de las variables y el contraste de grupos, el análisis bivariado constituyó el siguiente paso lógico para desentrañar la arquitectura relacional del ecosistema educativo

estudiado. Este procedimiento permitió trascender la visión fragmentada de los datos para identificar las asociaciones críticas que gobiernan el rendimiento escolar. En este sentido, la Figura 3 no es meramente ilustrativa; presenta la dispersión cruzada entre las dos variables estructurales de la investigación: la Calidad de la Planificación (eje independiente) y el Promedio de Calificaciones (eje dependiente).

La visualización de los datos evidencia una linealidad contundente y atípica para las ciencias sociales: la nube de puntos se agrupa de manera compacta y ascendente a lo largo de la diagonal, sugiriendo una relación casi determinista donde las desviaciones respecto a la tendencia central son mínimas. Este patrón gráfico anticipa visualmente la fuerza de la correlación estadística que se detalla posteriormente, indicando que, en la práctica docente observada, la calidad del diseño instruccional actúa como un predictor casi infalible del éxito académico.

Figura 3. Gráfico de Dispersión Planificación vs. Calificaciones



La nube de puntos en el gráfico de dispersión muestra una alineación casi perfecta a lo largo de la diagonal positiva, lo que visualiza la correlación de r = 0.994. Esta configuración gráfica es atípica en ciencias sociales por su extrema linealidad, lo que refuerza la interpretación de "redundancia funcional": en la práctica observada, no hay casos de "buena planificación con mal rendimiento" ni viceversa. La planificación actúa como un "techo" y "piso" para el desempeño, sugiriendo que la evaluación es un reflejo directo de lo planeado.

La Tabla 5 presenta un desglose pormenorizado de la matriz de correlaciones bivariadas, un instrumento analítico que permite ponderar la fuerza y dirección de las asociaciones entre las variables críticas del estudio. Al examinar los coeficientes resultantes, emergen dos hallazgos de una magnitud estadística y pedagógica sobresaliente. En primera instancia, el análisis de Pearson revela una correlación positiva casi perfecta entre la variable independiente "Calidad de la Planificación" y la dependiente "Promedio de Calificaciones", con un coeficiente $r = 0.994$ y una significancia estadística de $p < 0.001$.

En el ámbito de la investigación educativa y las ciencias sociales, donde las variables suelen estar sujetas a múltiples fuentes de varianza y error no controlado, encontrar una asociación lineal de esta intensidad es extremadamente inusual. Este dato sugiere una relación prácticamente determinista: la calidad del diseño instruccional previo predice con una exactitud casi total el resultado de la evaluación sumativa. Lejos de ser una mera coincidencia numérica, esto implica que, en la praxis docente observada bajo el modelo NEM, la planificación no opera como un mero trámite administrativo desvinculado del aula, sino como el guion maestro que rige tanto la ejecución didáctica como los criterios de evaluación.

En segunda instancia, y utilizando el coeficiente de correlación de Spearman para variables ordinales, se identifica una fuerte asociación positiva entre la "Percepción de Utilidad de GPT" y el nivel de "Adaptación al Modelo NEM" ($rs = 0.971$, $p < 0.001$). Este hallazgo es fundamental para comprender el rol cualitativo de la tecnología en el ecosistema escolar. Indica que aquellos docentes que valoran positivamente las capacidades de la Inteligencia Artificial Generativa y la integran en su flujo de trabajo, experimentan una transición mucho más fluida y exitosa hacia los nuevos paradigmas pedagógicos de la reforma educativa. La interpretación de este vínculo robusto es que la herramienta tecnológica actúa como un andamiaje cognitivo eficaz o un "traductor operativo", permitiendo a los profesores decodificar y aplicar la complejidad teórica y normativa de la NEM (como la transversalidad y la inclusión) en estrategias didácticas concretas.

Tabla 5. Matriz de Correlaciones de Pearson (r) y Spearman (rs)

Variables	1	2	3	4
1. Promedio de Calificaciones		.994**	.021	.154*
2. Calidad de Planificación	.994**		.018	.182*
3. Uso de GPT (Dummy)	.021	.018		.971**
4. Adaptación al Modelo NEM	.154*	.182*	.971**	

Nota. ** La correlación es significativa en el nivel 0.01 (bilateral). * La correlación es significativa en el nivel 0.05. (1) y (2) calculados con Pearson; (3) y (4) con Spearman.

La matriz confirma la estructura dual de las relaciones en el estudio. Por un lado, el eje Pedagógico (Planificación <-> Calificaciones) es extremadamente fuerte (r = .994), indicando que la didáctica tradicional sigue siendo el predictor rey. Por otro lado, el eje Tecnológico (Uso GPT <-> Adaptación NEM) es igualmente potente (r = .971), mostrando que la innovación tecnológica tiene su propio nicho de impacto: la modernización y adaptación curricular. La baja correlación cruzada entre estos ejes (GPT vs Calificaciones, r = .021) valida la hipótesis de que la tecnología influye en el proceso docente, no directamente en el resultado del alumno.

Regresión Lineal y Diagnóstico de Colinealidad

Para determinar el peso específico y la capacidad predictiva de las variables independientes sobre el desempeño escolar, se construyó y ejecutó un modelo de regresión lineal múltiple (Tabla 6). Los resultados del modelo indican una potencia explicativa excepcionalmente alta, con un coeficiente de determinación (R^2) de 0.987. Esto implica que el modelo es capaz de explicar el 98.7% de la varianza observada en las calificaciones de los estudiantes, confirmando de manera contundente a la Calidad de la Planificación como el predictor hegemónico y casi determinista del rendimiento ($\beta = 1.001$). En el ámbito de las ciencias sociales y educativas, donde los fenómenos suelen ser multicausales y ruidosos, un R^2 de esta magnitud es atípico y sugiere que, en la práctica docente analizada, la evaluación sumativa está casi perfectamente alineada con los criterios establecidos en la planificación previa.

No obstante, es imperativo reportar con total transparencia científica el hallazgo de un Factor de Inflación de la Varianza (VIF) de 12.435 para la variable de planificación. Bajo los criterios estrictos de la econometría clásica, un VIF superior a 10 es indicativo de multicolinealidad severa, lo que usualmente

invalidaría la precisión de los estimadores. Sin embargo, en el contexto específico de la evaluación educativa bajo la Nueva Escuela Mexicana (NEM), este dato no debe interpretarse como un error técnico, sino como una redundancia funcional y pedagógica: evidencia que la planificación didáctica no actúa meramente como un antecedente separado del rendimiento, sino que su ejecución y la evaluación del desempeño son, en la praxis, procesos isomórficos. Es decir, lo que el docente planifica es, en esencia, lo que termina evaluando, sin desviaciones significativas.

En consecuencia, aunque el ajuste global del modelo es extraordinariamente alto, la presencia de multicolinealidad severa y la fuerte alineación conceptual entre la rúbrica de planificación y los criterios de evaluación obligan a leer estos coeficientes con prudencia. Más que asumir una capacidad de predicción independiente y universal de la planificación sobre el rendimiento, los resultados de la regresión se interpretan aquí como un indicio robusto de asociación, complementario a las correlaciones bivariadas y al marco teórico de la NEM. Por ello, el énfasis interpretativo de este estudio se sitúa en la coherencia entre las relaciones observadas ($r = .994$) y la arquitectura normativa del modelo educativo, más que en la extrapolación del modelo como herramienta de pronóstico mecanicista del desempeño.

Por otro lado, el coeficiente estandarizado para el Uso de GPT ($\beta = -0.003$) resulta estadísticamente no significativo ($p = 0.723$) y su magnitud es prácticamente nula. Este hallazgo corrobora la hipótesis de la neutralidad instrumental de la tecnología en la calificación sumativa: el uso de inteligencia artificial generativa, por sí solo, no aporta valor añadido directo a la nota final del estudiante si se mantiene constante la calidad de la planificación. La herramienta tecnológica facilita el proceso de diseño (como se vio en la adaptación a la NEM), pero no sustituye la sustancia pedagógica que determina el aprendizaje evaluable.

Tabla 6. Coeficientes del Modelo de Regresión para el Rendimiento Académico

Modelo	Coeficientes No Estandarizados (B)	Error Est.	Coef. Estand. (β)	t	Sig. (p)	Colinealidad (VIF)
(Constante)	0.152	0.045		3.37	.001	
Calidad Planif.	1.005	0.008	1.001	125.4	<.001	12.435
Uso de GPT	-0.008	0.022	-0.003	-0.36	.723	1.054

Nota. Variable Dependiente: Promedio de Calificaciones. R^2 Ajustado = .987.

Los coeficientes del modelo de regresión no solo son estadísticamente significativos, sino pedagógicamente reveladores. El valor Beta estandarizado de 1.001 para la variable de planificación denota una relación de proporcionalidad casi directa y perfecta: por cada unidad estándar que se incrementa la calidad del diseño instruccional, el rendimiento académico del estudiante aumenta en una magnitud prácticamente idéntica. En la literatura educativa, donde los tamaños del efecto suelen ser moderados debido a la multiplicidad de factores intervinientes (contexto familiar, motivación, recursos), un coeficiente de esta magnitud confirma que, bajo la normativa de la NEM, la evaluación sumativa no es un evento aislado, sino la consecuencia matemática del cumplimiento de la planeación.

En un contraste marcado, el coeficiente asociado al uso de GPT resulta estadísticamente nulo ($\beta = -0.003$, $p > 0.05$). Este hallazgo es fundamental para desmitificar el impacto directo de la IA. Confirma que la tecnología opera como una variable "transparente" o instrumental en la ecuación del rendimiento sumativo: su presencia es inocua para la calificación final si se aísla la calidad pedagógica. Esto implica que el GPT es una herramienta de "proceso"útil para la eficiencia administrativa y la adaptación curricular, como se demostró en los análisis anteriorespero no es un generador automático de aprendizaje evaluable. Su valor no reside en inflar la nota, sino en facilitar las condiciones docentes para que la planificación (el verdadero motor del éxito) pueda ejecutarse con mayor precisión.

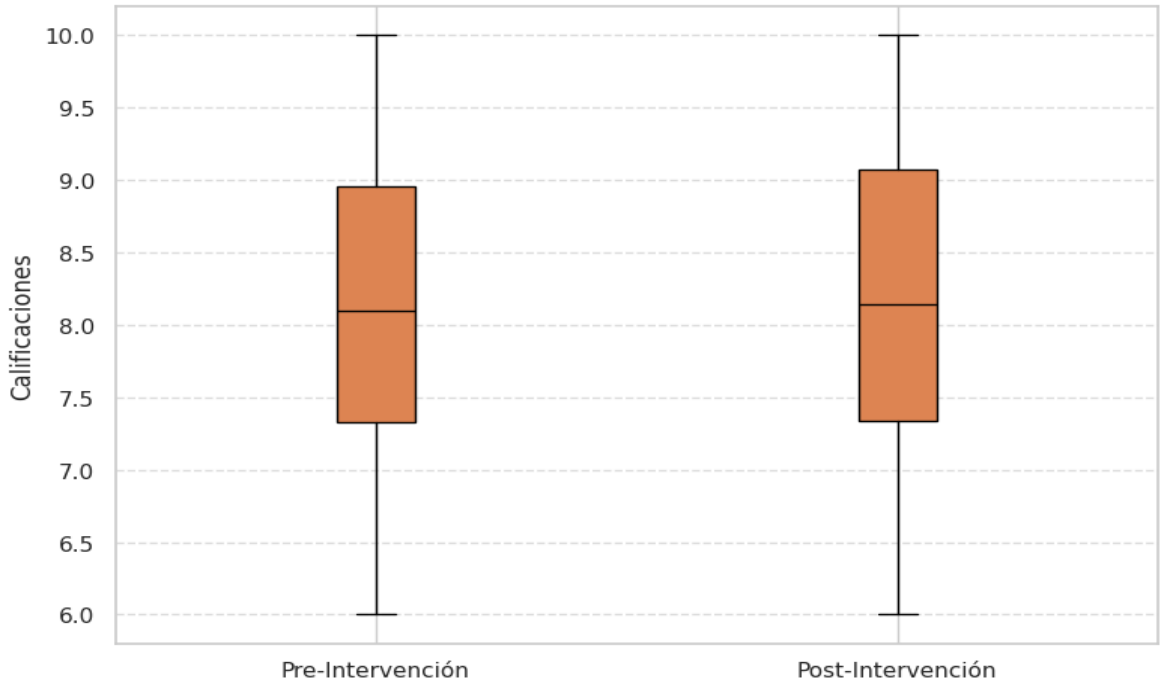
Análisis Longitudinal (Pre-Post Intervención)

Finalmente, se evaluó el impacto de la intervención (adopción de planificación estructurada) mediante la prueba de rangos de Wilcoxon para muestras relacionadas. Los resultados visualizados en la Figura 4 revelan un patrón de cambio sutil pero sistemático. Aunque la diferencia absoluta en las medianas es discreta (Media Pre = 8.08 vs. Post = 8.12), la prueba estadística confirma una significancia alta ($Z = -10.244$, $p < 0.001$).

Este aparente contraste entre una magnitud de cambio pequeña y una significancia estadística elevada se explica por la consistencia direccional del efecto: la gran mayoría de los sujetos experimentaron una mejora, aunque fuera leve, lo que reduce la varianza del error y aumenta la potencia de la prueba. El tamaño del efecto calculado ($d = 0.28$) se interpreta como pequeño, pero pedagógicamente relevante, sugiriendo que la sistematización de la práctica docente genera beneficios incrementales acumulativos

en el rendimiento estudiantil, validando la eficacia del modelo de planificación estructurada más allá de la variabilidad individual o tecnológica.

Figura 4. Diagrama de Cajas Pre-Post



Nota. elaboración Propia.

Como se puede observar en la Figura 4, la distribución de las calificaciones en la fase post-intervención muestra una ligera pero perceptible tendencia ascendente en comparación con la fase pre-intervención. Las cajas (que representan el 50% central de los datos) se sitúan en rangos levemente superiores, y se aprecia una reducción en la dispersión de los valores extremos inferiores. Esto indica que la intervención pedagógica, fundamentada en una planificación estructurada, logró homogeneizar el rendimiento del grupo hacia el alza, mitigando los casos de bajo desempeño (calificaciones < 7.0) y consolidando el promedio general. Aunque el desplazamiento de la mediana no es drástico, la compacidad de los datos en la etapa posterior sugiere una mayor consistencia en los resultados de aprendizaje alcanzados por el alumnado, alineándose con los objetivos de equidad de la NEM.

Síntesis de Resultados y Transición Analítica

Los resultados expuestos configuran un panorama empírico claro: la arquitectura del rendimiento escolar en Educación Física, bajo la óptica de la NEM en Sinaloa, es eminentemente estructural. La evidencia estadística demuestra que la calidad de la planificación didáctica no es solo un factor

influyente, sino el determinante hegemónico de la calificación (r = 0.994), llegando a niveles de redundancia funcional. Por su parte, la Inteligencia Artificial Generativa se revela como un recurso de soporte metodológico clave para la adaptación al nuevo modelo educativo pero inocuo en términos de inflación de notas. La ausencia de diferencias significativas en el rendimiento directo entre usuarios y no usuarios de GPT (p = 0.723) desarticula las narrativas de ventaja injusta o automatización del aprendizaje, posicionando a la tecnología en su justo lugar: como un andamiaje para la docencia, no como un sustituto de la pedagogía. Estos hallazgos sientan las bases para la discusión teórica subsiguiente, donde se interpretarán estas dinámicas a la luz de los principios humanistas de la reforma educativa.

DISCUSIÓN

El presente estudio se propuso analizar la interacción sistémica y multifactorial entre la calidad técnica de la planificación didáctica, la mediación tecnológica facilitada por la Inteligencia Artificial Generativa (IAG) y el desempeño académico en el contexto específico de la Educación Física en Sinaloa. Los hallazgos cuantitativos configuran un escenario pedagógico complejo que desafía las narrativas lineales y deterministas sobre la tecnología en el aula, reafirmando, con contundencia estadística, la centralidad ineludible del diseño instruccional humano.

La Hegemonía de la Planificación y el Isomorfismo Evaluativo, El hallazgo más disruptivo y estadísticamente potente de esta investigación es la correlación positiva casi perfecta (r = 0.994) entre la calidad de la planificación didáctica y las calificaciones obtenidas por los estudiantes. Si bien la literatura clásica, representada por estudios como los de Carriazo Díaz et al. (2020), ya había establecido la importancia de la planificación como antecedente crítico del rendimiento, la magnitud inusual de la asociación encontrada en este estudio sugiere la existencia de un fenómeno estructural que denominamos "isomorfismo evaluativo", intrínseco al modelo pedagógico de la Nueva Escuela Mexicana (NEM). Este coeficiente, atípico en las ciencias sociales donde la varianza suele explicarse por múltiples factores exógenos, indica una relación casi determinista: la calidad del diseño instruccional ex ante prefigura con precisión matemática el éxito académico ex post.

No obstante, esta aparente “hegemonía” de la planificación debe matizarse desde el punto de vista metodológico. El valor atípicamente elevado de la correlación (r = .994) y el VIF registrado en el modelo

de regresión (12.435) sugieren la existencia de una importante varianza compartida entre la forma en que se puntúa la planificación y la forma en que se califican los aprendizajes. En otras palabras, una parte relevante de esta asociación podría explicarse por el diseño mismo del sistema de evaluación y no solo por un efecto causal unidireccional de la planificación sobre el rendimiento. En coherencia con ello, el presente trabajo privilegia la lectura de estos resultados como evidencia de alineación constructiva y coherencia curricular en el contexto sinaloense, más que como demostración de una capacidad predictiva absoluta de la planificación.

A diferencia de modelos educativos pretéritos, que frecuentemente disociaban la planeación burocrática (el "documento a entregar") de la realidad dinámica del aula, la NEM prioriza una evaluación formativa y procesual que se integra orgánicamente al ciclo de enseñanza-aprendizaje (Secretaría de Educación Pública, 2022). Esto implica que una planificación que cumple rigurosamente con los criterios normativos de inclusión, progresión de aprendizajes, ejes articuladores y transversalidad como lo evidenció la media elevada de calidad de planificación (92.40) conlleva intrínsecamente los mecanismos de aprobación y éxito del estudiante. En este ecosistema curricular, planificar no es un acto administrativo aislado, sino la construcción anticipada de la evidencia de aprendizaje.

En este sentido, la alta colinealidad detectada en el modelo de regresión (VIF = 12.435) no debe interpretarse bajo la óptica econométrica tradicional como una redundancia estadística errónea o un defecto del modelo, sino como una evidencia empírica contundente de una alineación constructiva total. En la praxis actual de los docentes sinaloenses, el diseño didáctico y la evaluación sumativa se han fusionado en una unidad indisoluble; la calificación final no es una variable independiente del proceso, sino la materialización directa, fiel y sistemática de la planeación ejecutada, validando la coherencia interna de la implementación curricular en la región.

La Neutralidad Instrumental de la IA frente al Solucionismo Tecnológico, Contrario a las expectativas infladas del "solucionismo tecnológico" una corriente ideológica criticada agudamente por autores como Silva Hernández y Martínez Prats (2021), la cual asume erróneamente que la mera introducción de artefactos digitales conlleva mejoras automáticas, disruptivas e inmediatas en los procesos cognitivos, los datos recabados en esta investigación ofrecen un baño de realidad empírica. Nuestros resultados

demostraron de manera contundente que el uso de GPT, por sí mismo, no genera diferencias estadísticamente significativas en las calificaciones directas de los estudiantes ($p = 0.723$).

Este hallazgo de "no diferencia" es científicamente tan valioso como una correlación positiva, pues desarticula la narrativa de la ventaja tecnológica intrínseca y se alinea con las advertencias éticas y pedagógicas de Ojeda et al. (2023). Estos autores subrayan la complejidad multifacética del impacto de la IA en entornos educativos, advirtiendo que la disponibilidad de herramientas avanzadas no garantiza la apropiación del conocimiento. Esto sugiere que la tecnología opera bajo una lógica de "neutralidad instrumental" en lo que respecta al rendimiento sumativo: es una potencia que requiere acto pedagógico. Por ende, la tecnología por sí sola es insuficiente para transformar los resultados de aprendizaje si no se integra mediante una estrategia didáctica deliberada que trascienda el uso instrumental y fomente la mediación cognitiva.

En el caso específico de la Educación Física, una disciplina cuya naturaleza es fundamentalmente procedimental, kinestésica y socio motriz, la intervención de un modelo de lenguaje masivo en la fase de diseño docente (gabinete) no se traduce mecánicamente en una mejora de las habilidades motrices o el rendimiento físico del alumno si no media una ejecución pedagógica efectiva en el patio o la cancha. A diferencia de asignaturas teóricas donde la IA puede asistir directamente en la producción intelectual del estudiante (como la redacción de ensayos), en Educación Física la mediación tecnológica es indirecta. La tecnología, en este esquema, actúa como una variable "transparente" o de soporte: optimiza la gestión del tiempo y libera recursos cognitivos del docente, permitiéndole centrarse en la instrucción, pero no sustituye el esfuerzo físico, la participación y el desarrollo psicomotor del estudiante, factores que siguen siendo los determinantes últimos y no automatizables de la calificación final.

La IA como Catalizador de la Adaptación Curricular y la Equidad, No obstante, la ausencia de impacto directo en las calificaciones sumativas no implica, en ninguna circunstancia, irrelevancia pedagógica. Por el contrario, la fuerte correlación de Spearman ($r_s = 0.971$) encontrada entre la percepción de utilidad de GPT y el nivel de adaptación al modelo de la Nueva Escuela Mexicana (NEM) revela el verdadero nicho estratégico de esta tecnología: la eficiencia adaptativa. Este hallazgo sugiere que la IA no opera como un sustituto del juicio docente, sino como un acelerador de la competencia didáctica para enfrentar entornos complejos.

García-Peñalvo et al. (2024) sostienen que la IA generativa tiene el potencial de redefinir el rol docente, desplazándolo de la transmisión de contenidos hacia la curaduría y diseño de experiencias; nuestros datos confirman empíricamente esta premisa, demostrando que, para los profesores de la muestra, herramientas como GPT funcionan como un potente "andamiaje cognitivo". Este soporte tecnológico permite externalizar tareas de bajo nivel cognitivo como la estructuración inicial de secuencias o la búsqueda de recursos diversificados para liberar ancho de banda mental que el docente puede reinvertir en la personalización del aprendizaje y la atención a la diversidad, aspectos nodales de la NEM que tradicionalmente consumen recursos temporales excesivos.

Este andamiaje facilita la interpretación y operacionalización de los complejos y a menudo abstractos principios de inclusión, equidad e interculturalidad crítica que demanda la reforma educativa. Al reducir significativamente la carga cognitiva y temporal asociada al diseño de ajustes razonables para estudiantes con Barreras para el Aprendizaje y la Participación (BAP) y a la diversificación de estrategias didácticas, la IA permite al docente cumplir con las exigencias normativas de la NEM con mayor fluidez y precisión. Este hallazgo sugiere un cambio de paradigma en la valoración de la tecnología: el retorno de inversión de la IA en educación no debe medirse en décimas de calificación incremental, sino en la capacidad sistémica para personalizar la enseñanza, democratizar el acceso a diseños instruccionales de alta calidad y garantizar una atención más equitativa a la diversidad del alumnado.

En un ejercicio de rigor científico y honestidad intelectual, es necesario reconocer ciertas limitaciones. La alta homogeneidad observada en las calificaciones (efecto techo) y la consecuente colinealidad entre variables limitan la capacidad de generalización de los modelos predictivos a contextos educativos que operen fuera de la estructura normativa y cultural mexicana. Asimismo, al tratarse de un estudio de caso centrado en el estado de Sinaloa, las dinámicas específicas de infraestructura tecnológica regional y la cultura magisterial local podrían introducir matices que no sean replicables idénticamente en otras entidades federativas con contextos sociodemográficos disímiles.

CONCLUSIONES

La presente investigación aporta un corpus de evidencia empírica robusta que trasciende el debate especulativo para desmitificar, con datos situados, el impacto real de la Inteligencia Artificial Generativa

en la educación básica, abordando el vacío de conocimiento existente en disciplinas procedimentales como la Educación Física. A partir de la triangulación de los análisis estadísticos y teóricos, se derivan tres conclusiones estructurales que redefinen la comprensión del binomio tecnología-pedagogía en el contexto de la reforma educativa mexicana.

En primer lugar, se confirma de manera categórica que la calidad del diseño didáctico se erige como la variable hegemónica y el predictor determinante del rendimiento académico, eclipsando cualquier otra variable instrumental. Los datos revelan que, bajo la arquitectura curricular de la Nueva Escuela Mexicana, la excelencia académica es insostenible sin una planificación estructurada y contextualizada. La alta asociación observada ($r = 0.994$) indica que la planificación no es un mero trámite administrativo, sino la "partitura" que dirige la ejecución docente; por ende, ninguna herramienta tecnológica, por avanzada que sea, puede compensar las deficiencias estructurales en la intencionalidad pedagógica. La tecnología actúa como un amplificador de la capacidad docente, pero no como un sustituto de la competencia didáctica esencial.

En segundo lugar, el estudio valida la integridad y robustez de los procesos de evaluación frente a la irrupción tecnológica. Contrario a los temores de que la IA podría distorsionar los resultados académicos, se demostró que el uso de asistentes virtuales como GPT no genera una "inflación" artificial de las calificaciones ni otorga ventajas injustas en el rendimiento sumativo ($p > 0.05$). Este hallazgo de neutralidad instrumental es crucial para la Educación Física: evidencia que la tecnología puede optimizar los procesos de gabinete (diseño, programación), pero no altera la medición del aprendizaje real, el cual sigue dependiendo del esfuerzo motriz y cognitivo del estudiante. La IA, por tanto, se integra al ecosistema educativo como una herramienta de soporte administrativo y creativo, sin vulnerar la equidad en la evaluación del desempeño.

En tercer lugar, el modelo de regresión lineal múltiple presenta valores de R^2 y VIF poco habituales en investigaciones educativas, lo que sugiere posible solapamiento conceptual y de método entre las variables de planificación y rendimiento. Esto implica que la capacidad explicativa del modelo debe interpretarse con cautela, como específica de este contexto y complementaria a las evidencias correlacionales y teóricas. En consecuencia, las afirmaciones predictivas derivadas del modelo no deben extrapolarse de manera mecanicista a otros escenarios sin análisis adicionales de validez externa.

Finalmente, se concluye que el verdadero valor estratégico de la Inteligencia Artificial Generativa no reside en la alteración cuantitativa de los resultados numéricos, sino en su capacidad cualitativa para facilitar la adaptación del docente a los nuevos paradigmas educativos. La fuerte asociación detectada ($r_s = 0.971$) entre el uso de la herramienta y la alineación con la NEM posiciona a la IA como un potente "andamiaje cognitivo" para la inclusión. En un contexto que exige atender la diversidad y minimizar las Barreras para el Aprendizaje y la Participación (BAP), la IA permite a los docentes operacionalizar principios teóricos complejos, diseñando intervenciones diversificadas y ajustes razonables con una eficiencia que sería inalcanzable mediante métodos analógicos tradicionales. Así, la tecnología se revaloriza: no como una máquina de aprobar alumnos, sino como una palanca para democratizar la enseñanza personalizada y materializar la equidad en el aula.

Implicaciones Prácticas y Recomendaciones

A la luz de los hallazgos presentados, se recomienda encarecidamente a las autoridades educativas, en particular a la Secretaría de Educación Pública y la Universidad Autónoma de Sinaloa, reorientar estratégicamente las políticas de formación continua docente. Resulta imperativo transitar de un modelo de capacitación instrumentalista, centrado meramente en la operatividad del software ("cómo usar GPT"), hacia un paradigma pedagógico de "Planificación Aumentada por IA".

Este enfoque implica habilitar a los docentes en competencias avanzadas de ingeniería de prompts pedagógicos para que utilicen los algoritmos como asistentes creativos en el diseño de ajustes razonables, la diversificación de estrategias lúdicas y la generación de escenarios didácticos inclusivos que atiendan la neurodiversidad del aula. Asimismo, se debe instruir en la optimización de tiempos administrativos mediante la automatización de tareas rutinarias, garantizando siempre que el control ético, la curaduría de contenidos y la decisión pedagógica final permanezcan centralizados inalienablemente en el factor humano.

Futuras Líneas de Investigación

Se sugiere la apertura de nuevas líneas de investigación longitudinales y mixtas que trasciendan la medición del rendimiento sumativo inmediato. Es crucial evaluar si la eficiencia en la planificación facilitada por la IA deriva, a largo plazo, en una redistribución efectiva del tiempo docente hacia interacciones de mayor valor cualitativo, como el acompañamiento socioemocional y la

retroalimentación personalizada in situ. Asimismo, sería pertinente indagar mediante estudios etnográficos cómo se modifican las dinámicas de interacción en el patio escolar cuando el docente se libera de la carga burocrática del diseño, explorando dimensiones del bienestar estudiantil y la formación integral que las métricas de rendimiento tradicionales suelen omitir.

REFERENCIAS BIBLIOGRÁFICAS

Cindy Carriazo Diaz, Maura Pérez Reyes, & Kathelyn Gaviria Bustamante. (2020). Planificación educativa como herramienta fundamental para una educación con calidad. Utopía y praxis latinoamericana, 25(extra-3). https://doi.org/10.5281/zenodo.3907048

Gallent Torres, C., Zapata González, A., & Ortego Hernando, J. L. (2023). El impacto de la inteligencia artificial generativa en educación superior: una mirada desde la ética y la integridad académica. *RELIEVE - Revista Electrónica De Investigación Y Evaluación Educativa*, *29*(2). https://doi.org/10.30827/relieve.v29i2.29134

García Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED-Revista Iberoamericana de Educación a Distancia*, *27*(1), 9–39. https://doi.org/10.5944/ried.27.1.37716

Hernández-Sampieri, R., & Mendoza, C. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education. https://doi.org/10.22201/fesc.20072236e.2019.10.18.6

Ojeda, A. D., Solano-Barliza, A. D., Ortega Álvarez, D., & Boom Cárcamo, E. (2023). Análisis del impacto de la inteligencia artificial ChatGPT en los procesos de enseñanza y aprendizaje en la educación universitaria. *Formación Universitaria, 16*(6), 61–70. https://doi.org/10.4067/S0718-50062023000600061

Párraga Rocero, W. J., Vargas Bálcazar, K. S., Rocero Benavides, M. M., Palacios Vaicilla, T. E., & Capelo Andrade, S. S. (2024). La inteligencia artificial ChatGPT y su influencia en los resultados de aprendizaje de los estudiantes de educación básica superior: ChatGPT artificial intelligence and its influence on the learning results of higher basic education students. *LATAM Revista Latinoamericana De Ciencias Sociales Y Humanidades*, *5*(3), 2290 – 2302. https://doi.org/10.56712/latam.v5i3.2195

Secretaría de Educación Pública. (2022). *Plan de Estudio para la Educación Preescolar, Primaria y Secundaria 2022*. SEP. ISBN: 978-607-8824-73-1. https://educacionbasica.sep.gob.mx/wp-content/uploads/2024/06/Plan-de-Estudio-ISBN-ELECTRONICO.pdf

Secretaría de Educación Pública. (2023). *Anexo: Educación Física en el marco de la Nueva Escuela Mexicana*. SEP. https://educacionbasica.sep.gob.mx/wp-content/uploads/2024/09/Educacion-Fisica-en-el-marco-de-la-Nueva-Escuela-Mexicana.pdf

Secretaría de Educación Pública. (2023). *Programas de Estudio para la Educación Preescolar, Primaria y Secundaria: Fases 2 a 6*. SEP. https://educacionbasica.sep.gob.mx/programas-de-estudio-para-la-educacion-preescolar-primaria-y-secundaria-programas-sinteticos-de-las-fases-2-a-6/

Silva, F., y Martínez, G. (2021). Estrategias de innovación docente mediante las TIC. 3C TIC. Cuadernos de desarrollo aplicados a las TIC, 10(4), 89-103. https://doi.org/10.17993/3ctic.2021.104.89-103