International Education Trend Issues
Volume 4, Number 1, 2026
https://ijble.com/index.php/ieti

Artificial Intelligence in Educational Administration: A Systematic Evidence Review and Exploratory Meta-Analysis of Empirical Studies, 2020–2025

Matyoqubovich Sobirov
Bukhara State University, Uzbekistan
bobor@gov.co

Abstract
Artificial intelligence (AI) is increasingly used to automate administrative work, support institutional decision-making, optimise resources, and improve communication in educational organisations. Yet the empirical evidence remains fragmented, dominated by perception studies, and methodologically heterogeneous. This study synthesised empirical research published from 2020 to 2025 on AI-based educational administration and conducted an exploratory meta-analysis of statistically compatible outcomes. A structured search of scholarly indexes, publisher repositories, and citation networks identified 11 empirical studies covering school and higher-education settings. The narrative corpus represented more than 2,400 respondent- or school-level units and 150 administrative records. Only two studies (combined N = 260) reported sufficiently compatible statistics for standardised effect-size estimation. Using the author-reported pre–post effect, a random-effects model yielded Hedges' g = 1.00, 95% CI [0.24, 1.77], with substantial heterogeneity (I^2 = 90.2%). A sensitivity analysis reconstructing the pre–post effect from the reported t statistic produced a more conservative pooled estimate of g = 0.65, 95% CI [0.45, 0.86]. Narrative findings indicated potential improvements in processing time, reporting accuracy, workflow coordination, decision support, and resource allocation. However, most studies were cross-sectional, single-site, perception-based, or lacked control groups, and one effect estimate showed internal statistical inconsistency. The evidence therefore supports a cautiously positive conclusion: AI can enhance educational administration when it is embedded in redesigned workflows, staff development, human validation, and risk-based governance, but the current evidence base is insufficient for strong causal or universal claims. Future research should use controlled multisite designs, standardised administrative outcome measures, longitudinal evaluation, transparent system descriptions, cost-effectiveness analysis, and explicit auditing of privacy, fairness, explainability, and accountability.

Keywords: artificial intelligence; educational administration; educational management; school leadership; administrative efficiency; digital transformation; human–AI collaboration; meta-analysis

DOI. https://doi.org/10.56442/ieti.v4i1.1541

1. Introduction

Educational organisations depend on administrative systems that coordinate admissions, student records, staffing, timetables, finance, procurement, facilities, assessment reporting, communication, compliance, and institutional planning. Although many institutions have digitised these processes, conventional information systems generally store and retrieve data rather than interpret patterns, generate recommendations, or adapt workflows. Artificial intelligence extends digital administration by enabling prediction, classification, optimisation, natural-language processing, anomaly detection, and content generation (Ahmad et al., 2021, 2022; Chen et al., 2020; Hwang et al., 2020).

The rapid diffusion of generative AI has widened the range of administrative tasks that can be partially automated. School and university personnel can use large language models to draft correspondence, synthesise reports, prepare meeting materials, translate documents, generate policy options, and retrieve information from institutional knowledge bases. Predictive systems can support enrolment planning, resource forecasting, student-risk monitoring, and workload allocation. These capabilities make educational administration a strategically important but under-researched domain of AI in education (Crompton & Burke, 2023; Chiu et al., 2023; Sposato, 2025).

The dominant AI-in-education literature has focused on tutoring, assessment, personalisation, and learner analytics. Systematic reviews consistently show that administrative and leadership applications occupy a smaller share of the evidence base than teaching and learning applications, while educators and institutional decision-makers remain underrepresented in system design and evaluation (Bond et al., 2024; Mustafa et al., 2024; Zawacki-Richter et al., 2019). This imbalance matters because administrative efficiency affects the time, attention, and organisational capacity available for educational improvement.

AI adoption also creates risks. Administrative systems process sensitive information and can influence access, progression, staffing, discipline, and resource allocation. Inaccurate, biased, or opaque recommendations may therefore produce material consequences for students and employees. Human-centred research warns that stakeholder involvement, human control, safety, reliability, and trustworthiness are often treated superficially in educational AI projects (Alfredo et al., 2024). International governance frameworks similarly emphasise transparency, accountability, privacy, robustness, human oversight, and contestability (European Parliament and Council of the European Union, 2024; Miao & Holmes, 2023; National Institute of Standards and Technology [NIST], 2023; OECD, 2024; UNESCO, 2021).

The evidence problem is not a lack of optimistic claims but a lack of comparable outcome data. Studies use diverse definitions of AI, vary in implementation maturity, and frequently measure attitudes or perceived usefulness rather than objectively verified administrative performance. Cross-sectional associations may also reflect selection: institutions with stronger leadership, infrastructure, or digital capability may be more likely to adopt AI and may already perform better. Consequently, causal conclusions require more than positive user perceptions.

This review addresses that gap by integrating empirical studies of AI in educational administration published between 2020 and 2025. It combines a structured narrative synthesis with a deliberately conservative exploratory meta-analysis. The review asks: (RQ1) What administrative functions and outcomes have been examined empirically? (RQ2) What is the direction and magnitude of the available quantitative evidence for administrative efficiency? (RQ3) Which technological, organisational, human, and governance conditions appear to shape successful implementation? (RQ4) What methodological limitations should guide the next generation of educational-technology research?

2. Literature Review and Conceptual Framework

2.1 Defining AI-Based Educational Administration

In this article, AI-based educational administration refers to the use of computational systems capable of learning, prediction, classification, optimisation, recommendation, natural-language processing, or generative production to support the planning, organisation, delivery, monitoring, and evaluation of educational services. The definition distinguishes AI from basic digitisation: a database that stores records is not necessarily AI, whereas a system that detects anomalies, predicts demand, recommends a timetable, or generates a context-sensitive report performs an AI-enabled function.

Administrative applications can be grouped into six functional categories: (a) transactional automation, including document classification, record completion, and routine responses; (b) predictive analytics for enrolment, staffing, risk, and demand; (c) decision support that produces ranked alternatives or scenario forecasts; (d) optimisation of schedules, rooms, workloads, and resources; (e) natural-language interfaces for communication, retrieval, summarisation, and translation; and (f) monitoring and anomaly detection. These functions align with broader taxonomies of AI in education while shifting attention from learner-facing systems to organisational processes (Ahmad et al., 2022; Hwang et al., 2020; Ouyang & Jiao, 2021; Sposato, 2025).

2.2 Adoption, Acceptance, and Organisational Change

Technology adoption is not determined solely by technical performance. The Technology Acceptance Model proposes that perceived usefulness and perceived ease of use influence intention and actual use (Davis, 1989), while the Unified Theory of Acceptance and Use of Technology adds performance expectancy, effort expectancy, social influence, facilitating conditions, and moderating characteristics (Venkatesh et al., 2003). Diffusion theory further explains how adoption progresses through stages and user categories, and how compatibility, complexity, trialability, observability, and relative advantage shape uptake (Rogers, 2003).

These theories are relevant to educational administration because AI alters roles, authority, and professional routines. Tyson and Sauers (2021) found that school leaders did not merely select an AI system; they created organisational structures, communication processes, and implementation supports around it. Berkovich (2025) similarly interpreted school leaders' generative-AI use through diffusion stages, showing that adoption was uneven across tasks and leadership domains. The implication is that administrative performance depends on the joint configuration of technology, workflow, capability, and leadership rather than the software alone. This interpretation is consistent with school-leadership analyses that frame AI adoption as an organisational and ethical change challenge (Fullan et al., 2024).

2.3 Human–AI Complementarity

AI systems are strongest at high-volume pattern recognition, rapid retrieval, consistency, and generation of alternatives. Human administrators contribute contextual interpretation, ethical judgement, relational knowledge, legal authority, and accountability. Wang (2021) therefore conceptualised educational leadership as a symbiotic decision process in which human and artificial intelligence contribute different capabilities. Ouyang and Jiao's (2021) three-paradigm framework likewise supports a progression from AI-directed activity to AI-supported collaboration and ultimately human agency supported by AI.

This complementarity is especially important in high-consequence administration. An algorithm may identify an anomalous application or forecast staffing demand, but a responsible official must evaluate data quality, contextual exceptions, equity implications, and the legitimacy of the decision. Human oversight is not a ceremonial approval step; it requires access to evidence, authority to override the system, and documentation of the final rationale.

2.4 Governance, Ethics, and Risk

Responsible deployment requires risk proportionality. Low-risk applications such as document summarisation differ materially from systems that influence admissions, discipline, employee evaluation, or financial allocation. The NIST AI Risk Management Framework organises practice around governance, mapping, measurement, and management of risk (NIST, 2023). UNESCO's Recommendation on the Ethics of Artificial Intelligence and its generative-AI guidance foreground human rights, inclusion, transparency, privacy, and human capacity (Miao & Holmes, 2023; UNESCO, 2021). The OECD AI Principles emphasise human-centred values, robustness, transparency, and accountability, while the European Union AI Act establishes a risk-based regulatory architecture and identifies certain educational uses as high-risk (European Parliament and Council of the European Union, 2024; OECD, 2024).

Empirical governance remains immature. A survey of educational leaders reported substantial policy gaps and limited attention to student privacy and algorithmic transparency (Ghimire & Edwards, 2024). Reviews of generative AI also identify hallucination, bias, privacy, academic integrity, overreliance, and unequal access as recurrent concerns (Kasneci et al., 2023; Tlili et al., 2023). These risks justify a socio-technical evaluation model in which performance metrics are examined together with error pathways, user competence, procedural fairness, and contestability.

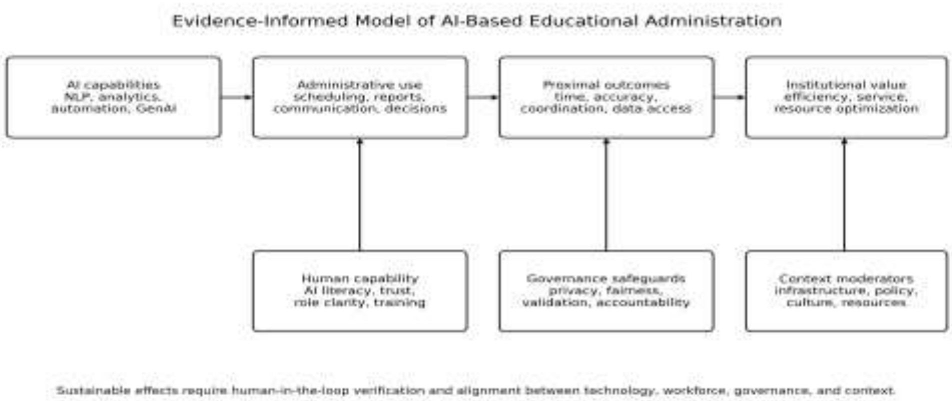

Figure 1. Evidence-informed conceptual model of AI-based educational administration. Sustainable value depends on the alignment of AI capabilities, administrative workflows, human capability, governance safeguards, and institutional context.

2.5 Research Gap and Analytical Propositions

The reviewed conceptual literature suggests three propositions. First, AI is most likely to produce measurable gains when it is applied to high-volume, rules-constrained, information-intensive tasks. Second, observed effects will be moderated by organisational readiness, including training, data quality, infrastructure, leadership, and workflow redesign. Third, higher levels of automation increase the need for explainability, validation, auditability, and appeal mechanisms. The empirical review evaluates whether the available evidence supports these propositions and whether reported effect sizes are sufficiently comparable for quantitative synthesis.

3. Method

3.1 Review Design and Reporting Standard

The study used a systematic evidence-review design with an exploratory meta-analysis. PRISMA 2020 informed the specification of eligibility criteria, search concepts, study selection, data extraction, and transparent reporting (Page et al., 2021). The review is described as a systematic evidence review rather than a fully database-audited systematic review because the accessible discovery interfaces did not provide a stable export of complete identification counts across all search sessions. No identification total was reconstructed or invented. This constraint is reported explicitly and should be addressed through a final Scopus, Web of Science, and ERIC export before journal submission.

The quantitative synthesis was prospectively restricted to outcomes that represented administrative efficiency or closely comparable operational performance and provided sufficient information to calculate a standardised mean difference. Because very few eligible studies reported compatible statistics, the meta-analysis was considered hypothesis-generating rather than confirmatory.

3.2 Search Strategy

Searches were completed on 30 July 2026 and restricted to studies published from 1 January 2020 to 31 December 2025. Sources included ERIC, Google Scholar discovery, Crossref-indexed publisher pages, journal repositories, and backward and forward citation chaining. The core search combined terms for AI, educational organisations, administrative functions, and empirical evaluation.

Core search string: ("artificial intelligence" OR "machine learning" OR "generative AI" OR "natural language processing") AND ("educational administration" OR "educational management" OR "school leadership" OR "academic leadership" OR "administrative efficiency" OR "administrative performance" OR "decision support") AND (school OR university OR education).

Table 1. Eligibility framework and operational scope

Element	Operational definition
Population/context	Schools, school systems, colleges, universities, and other formal educational organisations
Intervention/exposure	Use, adoption, implementation, or evaluation of an AI-enabled administrative or leadership function
Outcomes	Efficiency, time, accuracy, decision support, resource allocation, communication, readiness, adoption, trust, governance, or implementation
Study designs	Quantitative, qualitative, mixed-method, case study, quasi-experimental, and comparative empirical designs
Publication window	2020–2025
Language/access	English-language records with sufficient methodological and results information for verification

3.3 Eligibility Criteria

Studies were included when they: (a) reported original empirical data; (b) investigated an AI-enabled administrative, management, or leadership function in a formal educational setting; (c) described the sample and method; (d) reported findings relevant to implementation, use, perception, or performance; and (e) were published within the specified period. Studies were excluded when they were conceptual papers, editorials, reviews, non-educational applications, purely instructional studies without an administrative component, duplicate reports, or records without sufficient methodological information.

3.4 Data Extraction and Evidence Appraisal

Extracted variables included publication year, country, educational level, design, sample, AI function, administrative domain, outcome measures, descriptive statistics, inferential statistics, effect-size information, and reported limitations. Evidence quality was appraised across eight dimensions: design appropriateness, sample adequacy, measurement validity, presence of a comparator or baseline, use of objective outcomes, completeness of statistical reporting, control of confounding, and transparency of the AI intervention.

Because the corpus combined qualitative, cross-sectional, comparative, and quasi-experimental studies, a single design-specific risk-of-bias instrument would have produced misleading comparability. The review therefore used a structured cross-design appraisal and classified overall concerns as lower, moderate, or higher. Thematic synthesis followed an iterative coding process consistent with Braun and Clarke (2006), focusing on administrative benefits, mechanisms, implementation conditions, and governance risks.

3.5 Effect-Size Calculation

For independent-group studies, the standardised mean difference was calculated from the difference between group means divided by the pooled standard deviation. Cohen's d was corrected for small-sample bias to obtain Hedges' g (Borenstein et al., 2009; Hedges & Olkin, 1985).

$$d = (M_1 - M_0) / S_p$$
$$S_p = \sqrt{[(n_1 - 1)S_1^2 + (n_0 - 1)S_0^2] / (n_1 + n_0 - 2)}$$
$$g = Jd, \text{ where } J = 1 - 3 / (4df - 1)$$

For the pre–post study, the primary analysis used the author-reported standardised effect after small-sample correction. Because the reported effect was inconsistent with the accompanying t statistic, a sensitivity estimate was reconstructed as $d^z = t/\sqrt{n}$. This approach avoids treating the two estimates as if they were equally certain and directly exposes the dependence of the pooled result on one reporting decision.

3.6 Statistical Synthesis

The primary pooled estimate used a DerSimonian–Laird random-effects model because the studies differed in design, country, organisational setting, intervention, and measurement (DerSimonian & Laird, 1986). A fixed-effect estimate was reported descriptively. Heterogeneity was evaluated using Cochran's Q, τ^2, and I^2, with I^2 interpreted as the proportion of observed variability not attributable to sampling error (Higgins et al., 2003).

Publication-bias tests and funnel plots were not conducted because only two independent effects were available. Such analyses would be statistically uninformative and visually misleading. The meta-analytic results were therefore interpreted together with study quality, sensitivity analysis, and the wider narrative evidence.

4. Results

4.1 Study Profile

Eleven empirical studies met the inclusion criteria. The studies covered the United States, Saudi Arabia, Cyprus, Indonesia, Israel, Nigeria, Egypt, and a 12-country school sample. Designs included qualitative case studies, cross-sectional surveys, a quasi-experimental pre–post intervention, a mixed-method single-site evaluation, and secondary analysis of international school data. The evidence base was recent: one study was published in 2021, three in 2024, and seven in 2025. No eligible empirical administrative study was identified for 2020, 2022, or 2023 under the stated criteria.

Table 2. Characteristics and appraisal of included empirical studies

Study	Context	Design/sample	Administrative focus	Principal finding	Concern
Tyson & Sauers (2021)	United States; schools	Qualitative case study; 7 leaders	Adoption and implementation structures	Organisational support, communication, and routines shaped implementation	Moderate
Abdelaziz & Abdelshafy (2024)	Saudi Arabia; public schools	Cross-sectional survey; n = 411	Planning, scheduling, reporting, decisions, communication	High perceived administrative-performance contribution	Moderate
Masnun et al. (2024)	Indonesia; educational administration	Pre–post intervention; n = 50	Human–AI collaboration, role redesign, training	Perceived efficiency and capability improved after six months	Higher

Study	Context	Design/sample	Administrative focus	Principal finding	Concern
Pohan & Pohan (2025)	Indonesia; one madrasah	Mixed-method case; 20 staff, 150 records	Grade recap and reporting automation	Large reductions in processing time and input errors	Higher
Okafor et al. (2025)	Nigeria; education administrators	Comparative survey; n = 210	Efficiency, decisions, resource optimisation	AI users reported higher efficiency than non-users	Moderate
Selmi Arrooqi & Alruqi (2025)	Saudi Arabia; academic leaders	Cross-sectional survey; n = 105	Attitudes, readiness, benefits, ethics	Perceived benefit and ethics predicted readiness	Moderate
Kafa (2025)	Cyprus; school leaders	Qualitative interviews; n = 43	Management-task use cases	Perceived value for data, scenarios, communication, and planning	Moderate
Jamaica & Tagbo (2025)	Nigeria; universities	Descriptive survey; n = 48	ML, NLP, scheduling, effectiveness	Positive perceived use; limited inferential evidence	Higher
Riegel et al. (2025)	United States; multi-state K–12	Cross-sectional survey; denominators varied	Knowledge, use, perceptions, concerns	Administrators reported more frequent use than teachers	Moderate–higher
Berkovich (2025)	Israel; school leaders	Cross-sectional survey; n = 302	Generative AI in leadership and management	Use concentrated in instructional leadership, communication, and reports	Moderate
Eryilmaz et al. (2025)	12 countries; 1,134 schools	Secondary analysis of ICILS 2023	Expected impact, opportunity, and concern	Optimism coexisted with integrity and workload concerns	Lower–moderate

Concern categories are cross-design judgements rather than scores from a single risk-of-bias instrument. "Higher" indicates important design or reporting limitations, not evidence of misconduct.

4.2 Distribution of Evidence across Administrative Domains

The evidence was unevenly distributed. Readiness, adoption, and perceived usefulness were examined most frequently. Decision support and communication were also common, especially in qualitative studies and generative-AI surveys. Objective accuracy and time outcomes were rare. Only the Pohan and Pohan (2025) case study reported direct operational changes in processing duration and data-entry errors, and only two studies provided statistics suitable for a common standardised efficiency effect.

The evidence map in Figure 2 distinguishes perception or qualitative evidence from quantitative or operational evidence. It shows that the field has moved faster in documenting attitudes and use cases than in testing whether AI changes auditable organisational outcomes.

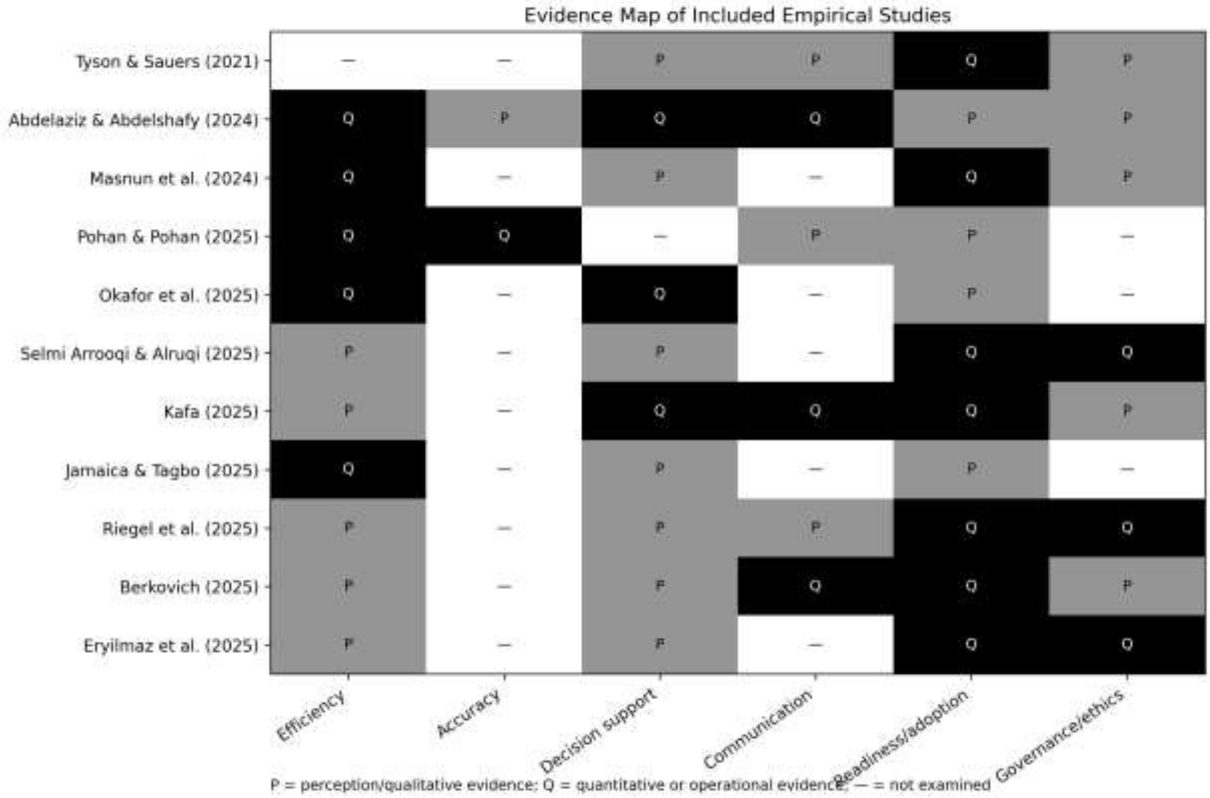

Figure 2. Evidence map of the included empirical studies. P indicates perception or qualitative evidence; Q indicates quantitative or operational evidence; a dash indicates that the domain was not examined.

4.3 Administrative Efficiency and Operational Performance

Okafor et al. (2025) compared 105 AI users with 105 non-users. The AI-user group reported higher administrative efficiency (M = 3.48, SD = 0.50) than the non-user group (M = 3.14, SD = 0.58), producing Hedges' g = 0.63, 95% CI [0.35, 0.90]. Secondary outcomes from the same sample also favoured AI users for resource optimisation (g = 0.41), data-supported decision-making (g = 0.44), and AI acceptance (g = 0.36). These secondary outcomes were not pooled because they were statistically dependent.

Masnun et al. (2024) evaluated a six-month intervention that combined AI use with role redefinition, training, and change management. Perceived efficiency increased from M = 4.05 (SD = 1.08) to M = 5.50 (SD = 0.95). The authors reported t(49) = 4.98 and Cohen's d = 1.43. After correction, the author-reported value corresponded to Hedges' g = 1.41. However, reconstruction from the t statistic yielded d^{z} = 0.70 and g = 0.69. This inconsistency materially affected the pooled estimate and was therefore treated as a central sensitivity issue rather than a minor reporting detail.

Pohan and Pohan (2025) reported that a single-site AI-supported management system reduced grade-recap processing from approximately two or three days to eight or ten minutes, reduced report preparation from three to five days to one or two hours, and decreased input errors by more than 90%. These results suggest a large operational opportunity when a well-defined manual workflow is automated. Nevertheless, the study lacked a concurrent comparator, variance estimates, independent verification, and multi-site replication; it could not be included in the standardised meta-analysis.

4.4 Adoption, Readiness, and Leadership Practice

Administrative adoption was generally positive but selective. Berkovich (2025) found that school leaders used generative AI more frequently for programme development, professional-learning materials, communication, and observation reports than for budgeting or resource allocation. This pattern is consistent with risk-sensitive adoption: text-intensive tasks are easier to trial and reverse than decisions affecting money, rights, or personnel.

Selmi Arrooqi and Alruqi (2025) reported high perceived benefits and found that ethical perceptions were associated with organisational readiness, while inadequate training remained a prominent barrier. Riegel et al. (2025) likewise observed more frequent AI use among administrators than teachers but documented persistent concerns about policy, privacy, support, and preparedness. Kafa (2025) found that school leaders anticipated value in data processing, scenario planning, decision support, and communication, while recognising the continuing need for human judgement.

International evidence suggested neither unqualified enthusiasm nor uniform resistance. Eryilmaz et al. (2025) found that leaders' expectations combined optimism about support and engagement with concern about integrity, workload, and implementation conditions. These mixed expectations are rational: the same system can reduce routine work while creating verification, governance, and professional-development burdens.

4.5 Exploratory Meta-Analysis

Table 3. Standardised effects for administrative efficiency

Study	Contrast	N	Hedges' g	95% CI	Fixed-effect weight
Okafor et al. (2025)	AI users vs non-users	210	0.63	0.35 to 0.90	66.6%
Masnun et al. (2024)	Pre–post; author-reported effect	50	1.41	1.02 to 1.80	33.4%
Fixed-effect pooled estimate		260	0.89	0.66 to 1.11	–
Random-effects pooled estimate		260	1.00	0.24 to 1.77	–

The primary random-effects analysis produced $g = 1.00$, $SE = 0.39$, 95% CI [0.24, 1.77]. Heterogeneity was substantial: $Q(1) = 10.23$, $\tau^2 = 0.276$, and $I^2 = 90.2\%$. The fixed-effect estimate was $g = 0.89$, 95% CI [0.66, 1.11]. Given the difference in design and the instability of heterogeneity estimates with two studies, the random-effects confidence interval is more informative than the point estimate alone.

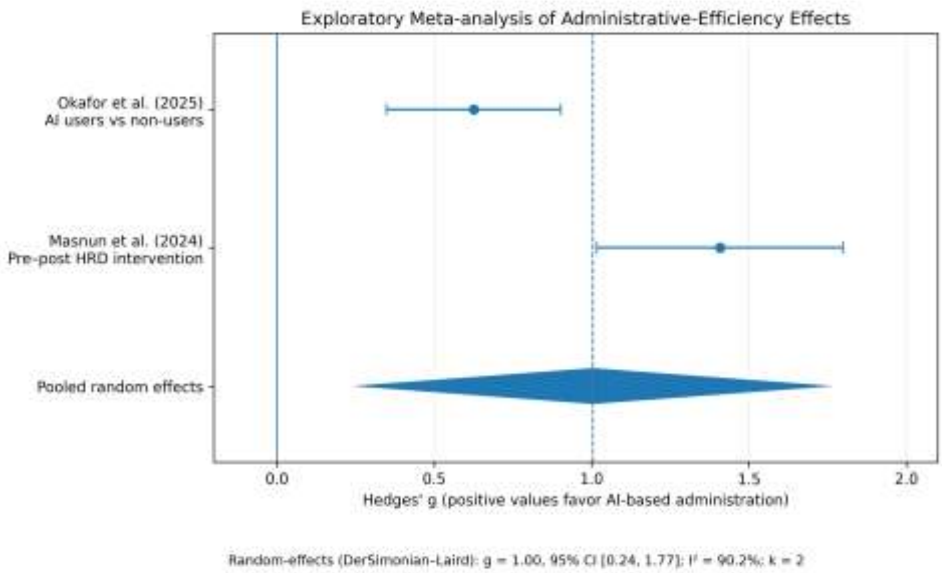

Figure 3. Exploratory forest plot for administrative-efficiency outcomes. Positive values favour AI-supported administration. The pooled diamond uses a DerSimonian–Laird random-effects model.

4.6 Sensitivity Analysis and Certainty of Evidence

Replacing the author-reported pre–post effect with the estimate reconstructed from $t/\sqrt{n}$ yielded a pooled effect of g = 0.65, 95% CI [0.45, 0.86], with Q(1) = 0.09 and I^2 = 0%. The contrast between the primary estimate (g = 1.00) and the sensitivity estimate (g = 0.65) demonstrates that the apparent magnitude and heterogeneity were driven largely by one internally inconsistent report.

The direction of effect remained positive in both scenarios, but the certainty of the pooled magnitude was low. Downgrading factors included very serious imprecision due to k = 2, indirectness caused by heterogeneous constructs and designs, risk of confounding, reliance on self-report, and possible publication bias that could not be tested. The appropriate conclusion is therefore that the available quantitative evidence is compatible with a moderate-to-large positive association, not that AI has a proven universal causal effect on educational administration.

Table 4. Certainty considerations for the quantitative conclusion

Domain	Observation	Judgement
Design and causality	Predominantly cross-sectional or uncontrolled	Serious
Measurement	Most outcomes were perceived usefulness or self-reported efficiency	Serious
Intervention transparency	AI tools and configurations were often incompletely described	Moderate to serious
Statistical reporting	One key pre–post study contained inconsistent effect information	Serious
Generalisability	Several studies were single-site or geographically concentrated	Serious
Consistency	Direction was generally positive, but comparable outcomes were scarce	Moderate
Publication bias	Not estimable with two quantitative effects	Unknown / serious concern

5. Discussion

5.1 Principal Findings

The synthesis supports a cautiously positive view of AI-based educational administration. Across diverse settings, AI was associated with faster processing, improved perceived efficiency, support for data-informed decisions, easier preparation of communication and reports, and more systematic resource planning. The quantitative evidence remained positive under both primary and sensitivity assumptions. The narrative evidence also converged on a practical pattern: AI generated the clearest benefits when applied to repetitive, information-intensive, and auditable tasks.

At the same time, the review does not support deterministic claims that AI inevitably improves educational management. The evidence base is small, recent, and dominated by self-report. Institutions that adopt AI may have stronger digital infrastructure, leadership, funding, or change capacity than non-adopters. Pre–post improvements may partly reflect training, attention, workflow redesign, or broader organisational development rather than the algorithm itself. The Masnun et al. (2024) intervention is a clear example: its package combined AI with human-resource development, so the estimated effect cannot be attributed to software alone.

5.2 Mechanisms of Administrative Value

Four mechanisms explain the observed benefits. First, automation reduces the labour required for repetitive classification, transcription, checking, summarisation, and response tasks. Second, integration and retrieval shorten the time needed to assemble evidence from fragmented records. Third, prediction and optimisation provide decision-makers with scenarios and ranked alternatives that would be difficult to generate manually. Fourth, generative systems lower the transaction cost of producing communication, reports, agendas, and policy drafts.

These mechanisms differ in evidential and ethical risk. Drafting a meeting summary is reversible and readily checked. Predicting student risk, ranking applicants, allocating budgets, or evaluating personnel can affect rights and opportunities. The expected efficiency gain must therefore be balanced against verification cost, false-positive and false-negative consequences, data-protection obligations, and the ability of affected parties to understand and challenge decisions.

5.3 Why Human Capability and Workflow Redesign Matter

The findings reinforce a socio-technical rather than tool-centric explanation of performance. AI value emerges when the task is well specified, data are fit for purpose, users understand the system, and the workflow defines how outputs are reviewed and acted upon. Training must therefore include more than prompt writing. Administrators need AI literacy, data literacy, domain validation skills, privacy awareness, bias recognition, and the authority to reject or escalate questionable outputs.

Workflow redesign is equally important. Adding AI to an inefficient process can automate waste or multiply errors. Institutions should first map the process, define the decision rights, establish a baseline, and identify failure modes. Only then should they automate components that are sufficiently structured and measurable. This sequence is consistent with the adoption evidence reported by Tyson and Sauers

(2021) and with broader human-centred design recommendations (Alfredo et al., 2024).

5.4 Governance Implications

Educational institutions require governance that is proportional to the consequence of the use case. At minimum, each AI application should have a named owner, documented purpose, authorised data sources, defined users, validation procedure, retention rules, incident process, and periodic review. High-consequence systems require additional controls: impact assessment, subgroup performance testing, explanation or reason-giving, independent audit where feasible, human approval, and an accessible appeal mechanism.

The governance approach can integrate international frameworks without treating them as interchangeable. NIST provides an operational risk-management cycle; UNESCO foregrounds human rights, inclusion, and capacity building; the OECD principles define characteristics of trustworthy AI; and the EU AI Act demonstrates how educational uses may be classified according to risk (European Parliament and Council of the European Union, 2024; Miao & Holmes, 2023; NIST, 2023; OECD, 2024; UNESCO, 2021). For educational administrators, the practical requirement is to translate these principles into controls that are visible in procurement, deployment, monitoring, and accountability.

Table 5. Evidence-informed implementation and evaluation framework

Stage	Required action	Minimum evidence or control
1. Problem diagnosis	Identify high-volume, delayed, costly, or error-prone processes	Process map; defined users and decision owner
2. Baseline measurement	Measure time, cost, backlog, error, satisfaction, and equity before AI	Auditable pre-implementation baseline
3. Risk classification	Assess data sensitivity, reversibility, affected rights, and severity of error	Documented low/medium/high-risk classification
4. Limited pilot	Test one bounded process with human review	Pilot accuracy, correction rate, and user feedback
5. Capability development	Train staff in use, verification, privacy, bias, and escalation	Competency assessment and role clarity
6. Comparative evaluation	Use before–after, matched comparison, or randomised rollout where feasible	Effect size, uncertainty, and cost-effectiveness
7. Governance and audit	Monitor drift, incidents, subgroup performance, complaints, and vendor changes	Audit trail, review schedule, appeal process
8. Scale or discontinue	Expand only when net value and risk controls are demonstrated	Decision based on evidence, not vendor claims

5.5 Implications for Educational Technology Research and Practice

Educational-technology research should broaden its unit of analysis from the learner–tool interaction to the institution as a coordinated socio-technical system. Administrative capacity can indirectly affect educational quality by releasing leadership and teacher time, improving information flow, reducing compliance burden, and enabling earlier intervention. Nevertheless, an administrative efficiency outcome should not be treated as educational value unless the saved resources are translated into better service, equity, or learning conditions.

A stronger evaluation agenda requires a core outcome set. Recommended operational outcomes include processing time, transactions per staff member, correction rate, backlog, cost per case, schedule quality, response time, decision turnaround, user workload, satisfaction, override frequency, appeal rate, privacy incidents, and subgroup error disparities. Researchers should also report the model or system, data provenance, level of automation, human-review process, implementation fidelity, training, vendor involvement, and changes made to the underlying workflow.

Economic evaluation is particularly important. The appropriate quantity is not gross time saved but net institutional value: operational benefit minus licence, integration, training, verification, governance, security, and error-remediation costs. An AI system that produces rapid drafts but requires extensive checking may shift rather than eliminate labour.

5.6 Limitations

a. Only two studies provided statistically compatible outcomes for pooling; heterogeneity and publication bias could not be estimated reliably.
b. The evidence was dominated by cross-sectional perceptions and uncontrolled designs, limiting causal inference.
c. Definitions of AI, implementation maturity, and administrative outcomes varied substantially across studies.
d. Several studies were single-site, had small samples, or were published in emerging journals with limited methodological detail.
e. One pre–post study reported an effect size that was inconsistent with its t statistic; the sensitivity analysis materially changed the pooled estimate.
f. The search relied on accessible scholarly discovery systems and publisher repositories. Stable identification counts were unavailable, so a conventional PRISMA flow with database-level record totals was not fabricated. A final submission should rerun and archive the search in Scopus, Web of Science, and ERIC.
g. The rapid evolution of generative AI means that systems, policies, and user practices may change faster than conventional publication cycles.

6. Conclusion

Empirical research published from 2020 to 2025 indicates that AI can improve selected educational-administration processes, particularly when tasks are repetitive, information-intensive, and amenable to verification. The exploratory meta-analysis produced a positive pooled association, but its magnitude ranged from moderate to large depending on the treatment of one inconsistent pre–post effect. The quantitative conclusion is therefore promising but low-certainty.

The most defensible policy conclusion is not that educational institutions should automate administration broadly, but that they should evaluate bounded use cases through measured pilots, human oversight, staff capability development, and risk-based governance. AI should augment professional judgement, not obscure responsibility. Stronger evidence will require controlled multisite studies, standardised operational outcomes, transparent system reporting, cost-effectiveness analysis, and explicit evaluation of privacy, fairness, explainability, and accountability.

References

Abdelaziz, G. A. R., & Abdelshafy, A. F. (2024). Using Artificial Intelligence Applications to Improve Administrative Performance in Public Education Schools. *Evolutionary Studies in Imaginative Culture*, *8.1*(S2), 2086-2099. https://doi.org/10.70082/esiculture.vi.2027

Ahmad, S. F., Rahmat, M. K., Mubarik, M. S., Alam, M. M., & Hyder, S. I. (2021). Artificial Intelligence and Its Role in Education. *Sustainability*, *13*(22), 12902. https://doi.org/10.3390/su132212902

Ahmad, S. F., Alam, M. M., Rahmat, M. K., Mubarik, M. S., & Hyder, S. I. (2022). Academic and Administrative Role of Artificial Intelligence in Education. *Sustainability*, *14*(3), 1101. https://doi.org/10.3390/su14031101

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-Centred Learning Analytics and AI in Education: A Systematic Literature Review. *Computers and Education: Artificial Intelligence*, *6*, 100215. https://doi.org/10.1016/j.caeai.2024.100215

Berkovich, I. (2025). The Rise of AI-Assisted Instructional Leadership: Empirical Survey of Generative AI Integration in School Leadership and Management Work. *Frontiers in Education*, *10*, 1643023. https://doi.org/10.3389/feduc.2025.1643023

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Phang, F. A., Chong, S. C., & Siemens, G. (2024). A Meta Systematic Review of Artificial Intelligence in Higher Education: A Call for Increased Ethics, Collaboration, and Rigour. *International Journal of Educational Technology in Higher Education*, *21*, 4. https://doi.org/10.1186/s41239-023-00436-z

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis.* John Wiley & Sons. https://doi.org/10.1002/9780470743386

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, *8*, 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic Literature Review on Opportunities, Challenges, and Future Research Recommendations of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Crompton, H., & Burke, D. (2023). Artificial Intelligence in Higher Education: The State of the Field. *International Journal of Educational Technology in Higher Education*, *20*, 22. https://doi.org/10.1186/s41239-023-00392-8

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319-340. https://doi.org/10.2307/249008

DerSimonian, R., & Laird, N. (1986). Meta-Analysis in Clinical Trials. *Controlled Clinical Trials*, *7*(3), 177-188. https://doi.org/10.1016/0197-2456(86)90046-2

Eryilmaz, N., Mah, D. K., Bellibaş, M. Ş., & Pietsch, M. (2025). School Leaders' Expectations of AI's Impact on Students and Teachers: Insights from ICILS 2023. *Educational Management Administration & Leadership*. https://doi.org/10.1177/17411432251401487

European Parliament and Council of the European Union (2024). *Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence*. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Fullan, M., Azorín, C., Harris, A., & Jones, M. (2024). Artificial Intelligence and School Leadership: Challenges, Opportunities and Implications. *School Leadership & Management*, *44*(4), 339-346. https://doi.org/10.1080/13632434.2023.2246856

Ghimire, A., & Edwards, J. (2024). *From Guidelines to Governance: A Study of AI Policies in Education*. arXiv. https://doi.org/10.48550/arXiv.2403.15601

Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring Inconsistency in Meta-Analyses. *BMJ*, *327*(7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, Challenges, Roles and Research Issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Jamaica, N. J., & Tagbo, S. (2025). Artificial Intelligence in Educational Management for Enhanced Administrative Effectiveness in Rivers State Universities. *International Journal of Educational Management, Rivers State University*, *1*(2), 213-236.

Kafa, A. (2025). Exploring School Leaders' Perceptions of Using an Artificial Intelligence Tool for Management Tasks: Empirical Evidence from Cyprus. *Management in Education*. https://doi.org/10.1177/08920206251394137

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Masnun, A., Jessika, S., Albadry, S. A., Zulkifli, & Hamirul (2024). Optimizing Human-AI Collaboration in Educational Administration in Muara Bungo, Jambi: An HRD Framework for Role Redefinition, Skill Development, and Change Management. *Enigma in Education*, *2*(2), 114-125. https://doi.org/10.61996/edu.v2i2.85

Miao, F., & Holmes, W. (2023). *Guidance for Generative AI in Education and Research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Mustafa, M. Y., Tlili, A., Lampropoulos, G., Huang, R., Jandrić, P., Zhao, J., Salha, S., Xu, L., Panda, S., Kinshuk, López-Pernas, S., & Saqr, M. (2024). A Systematic Review of Literature Reviews on Artificial Intelligence in Education (AIED): A Roadmap to a Future Research Agenda. *Smart Learning Environments*, *11*, 59. https://doi.org/10.1186/s40561-024-00350-5

National Institute of Standards and Technology (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Okafor, J. N., Oguejiofor, C. N., Ezenwagu, S. A., & Chibuko, H. (2025). Application of Artificial Intelligence in Educational Management. *NAEAP Journal of Studies in Educational Administration and Management*, *5*(1), 457-466.

Organisation for Economic Co-operation and Development (2024). *OECD AI Principles (Updated 2024).* OECD. https://oecd.ai/en/ai-principles

Ouyang, F., & Jiao, P. (2021). Artificial Intelligence in Education: The Three Paradigms. *Computers and Education: Artificial Intelligence*, *2*, 100020. https://doi.org/10.1016/j.caeai.2021.100020

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pohan, R., & Pohan, N. (2025). Integration of Artificial Intelligence in Islamic Education Management System to Improve Administrative Efficiency. *Sistem Pendukung Keputusan dengan Aplikasi*, 4(2), 101-110. https://doi.org/10.55537/spk.v4i2.1305

Riegel, C., Ford, A. Y., Brinkmann, J. L., Christian, G., Weinstein, D., & Cash, C. S. (2025). Exploring AI in Education: A Multi-State Study on K-12 Teachers' and Administrators' Knowledge, Use, and Perceptions of Artificial Intelligence. *Issues and Trends in Learning Technologies*, *13*(1).

Rogers, E. M. (2003). *Diffusion of Innovations.* (5th ed.). Free Press.

Selmi Arrooqi, A., & Alruqi, M. M. (2025). Academic Leadership Attitudes toward Employing Artificial Intelligence Applications in Developing Administrative Processes. *Humanities and Social Sciences Communications*, *12*, 1342. https://doi.org/10.1057/s41599-025-05598-x

Sposato, M. (2025). Artificial Intelligence in Educational Leadership: A Comprehensive Taxonomy and Future Directions. *International Journal of Educational Technology in Higher Education*, *22*, 20. https://doi.org/10.1186/s41239-025-00517-1

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What If the Devil Is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education. *Smart Learning Environments*, *10*, 15. https://doi.org/10.1186/s40561-023-00237-x

Tyson, M. M., & Sauers, N. J. (2021). School Leaders' Adoption and Implementation of Artificial Intelligence. *Journal of Educational Administration*, *59*(3), 271-285. https://doi.org/10.1108/JEA-10-2020-0221

UNESCO (2021). *Recommendation on the Ethics of Artificial Intelligence.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425-478. https://doi.org/10.2307/30036540

Wang, Y. (2021). Artificial Intelligence in Educational Leadership: A Symbiotic Role of Human-Artificial Intelligence Decision-Making. *Journal of Educational Administration*, *59*(3), 256-270. https://doi.org/10.1108/JEA-10-2020-0216

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic Review of Research on Artificial Intelligence Applications in Higher Education—Where Are the Educators?. *International Journal of Educational Technology in Higher Education*, *16*, 39. https://doi.org/10.1186/s41239-019-0171-0