© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722515

From Software Acquisition to Operational Capability: Measuring ERP-Enabled Performance

NKOSANA MKANDLA[1], MARLON MUNJOMA[2], MELODY SYDNEY[3], FLORA PHIRI[4], MUNASHE NAPHTALI MUPA[5]

[1]*University of Suffolk (UK), s286718@uos.ac.uk ORCiD: 0009-0001-8282-9714*
[2]*Pace University, mm28038n@pace.edu ORCiD: 0009-0004-7940-2172*
[3]*George Washington University, melodys@gwu.edu ORCiD: 0009-0003-4795-6755*
[4]*George Washington University, flora.phiri@gwu.edu ORCiD: 0009-0007-6433-3428*
[5]*Hult International Business School, ma1@student.hult.edu ORCiD: 0000-0003-3509-861X*

Abstract- Enterprise resource planning (ERP) projects are often evaluated at acquisition, configuration or go-live, although the economic and operational value of an integrated system depends on whether the organisation redesigns processes, governs data, embeds controls, develops user capability and measures realised benefits. This study develops an acquisition-to-capability framework and tests its operational measurement layer using a public 70,000-record analytical extract of the DataCo Smart Supply Chain dataset. The extract contains 44,026 unique orders dated from 2015 to 2018. Analysis was conducted at order and order-line levels using descriptive statistics, cross-segment heat maps, robust logistic regression, holdout prediction, calibration assessment, bootstrap confidence intervals and sensitivity testing of a six-component ERP Operational Capability Index (EOCI). The order-level late-delivery rate was 54.7%, severe delay affected 23.6%, negative-profit exposure affected 19.6%, and hard workflow exceptions affected 10.8%. Shipping configuration was the dominant observable delivery factor: compared with Standard Class, adjusted late-delivery odds were 34.22 times higher for First Class, 5.42 times higher for Second Class and 1.40 times higher for Same Day. A leakage-controlled holdout model achieved an area under the receiver operating characteristic curve of 0.731, a Brier score of 0.197, a calibration intercept of 0.008 and a calibration slope of 0.967. Observed on-time delivery rose from 11.2% in the lowest readiness quartile to 62.9% in the highest, a 51.6-percentage-point difference. EOCI rankings were stable under alternative weights (Spearman rho=0.998). The shipping-mode odds ratios reflect this dataset's own service-level design and are reported as a demonstration of what transaction-level analytics can reveal, not as a general finding about ERP-enabled delivery performance. The findings show that transaction visibility can diagnose capability gaps, but software records alone cannot establish implementation causality. The paper therefore specifies a matched difference-in-differences field protocol, KPI dictionary, governance gates and implementation playbook for SMEs and small government vendors. The central conclusion is that ERP value is realised through a governed operating system, not through software possession.

Keywords: enterprise resource planning; operational capability; process redesign; data governance; internal controls; user adoption; SME digital transformation; government vendors; supply-chain analytics; benefits realization

I. INTRODUCTION

Enterprise resource planning systems are purchased as integrated technologies, but their value is produced through organisational change. The distinction is consequential. A firm may own a capable platform while continuing to execute fragmented processes, maintain inconsistent master data, rely on spreadsheet workarounds, tolerate unresolved control exceptions and train users only in screen navigation. In that state, the organisation has acquired software without acquiring operational capability. Davenport (1998) described enterprise systems as technologies that impose process logic across organisational boundaries; subsequent research has shown that performance gains depend on complementary changes in processes, governance and use rather than on installation alone (Hitt, Wu and Zhou, 2002; McAfee, 2002; Hendricks, Singhal and Stratman, 2007).

The acquisition-capability gap is particularly important for small and medium-sized enterprises (SMEs) and small government vendors. These organisations face resource constraints, uneven digital skills and high dependence on key individuals, while simultaneously being expected to maintain accurate records, reliable delivery, timely invoicing and auditable controls. Cloud and modular ERP products have reduced entry barriers, but lower acquisition cost does not eliminate implementation risk. SMEs may digitise a weak process rather than redesign it, migrate poor-quality data, over-customise workflows or declare success at go-live before adoption and benefits are stable (Ruivo, Oliveira and Neto, 2014; Eller et al., 2020; AlMuhayfith and Shaiti, 2020).

The research problem is therefore not whether ERP can improve performance in principle. Evidence already links enterprise systems with changes in productivity, lead time, process agility, accounting visibility and organisational performance, although effects vary by implementation quality, organisational fit, timing and post-implementation learning (Nicolaou, 2004; Cotteleer and Bendoly, 2006; Gattiker and Goodhue, 2005; Seethamraju, 2013). The unresolved practical problem is how to measure the transition from software acquisition to operational capability in a way that is rigorous enough for research and usable enough for implementation teams.

This paper addresses that problem through two linked contributions. First, it develops a multi-domain measurement architecture connecting process redesign, data governance, controls, training, implementation governance and adoption with operational outcomes. Second, it demonstrates the transaction-level measurement layer using a public extract of the DataCo Smart Supply Chain dataset. The empirical analysis focuses on cycle-time conformance, delivery reliability, margin exposure, workflow exceptions, data completeness and a composite capability index. Because the public dataset lacks firm-level implementation dates and direct training or governance measures, the study does not attribute observed outcomes causally to ERP. Instead, it uses the data to establish benchmark diagnostics and then specifies a prospective matched-case protocol capable of estimating implementation effects.

This separation between measurement and causal attribution is deliberate. Contemporary analytics can produce persuasive dashboards while still failing to establish whether a technology intervention caused the observed change. A credible evaluation requires baseline data, a defined intervention, comparable controls, pre-trend assessment, adoption evidence and governance over metric definitions. The paper therefore treats operational analytics as one layer of a broader evidence system. This approach aligns with recent applied work on workflow optimisation, continuous ERP controls, decision intelligence and human-centred training analytics, which emphasises recurring monitoring, transparent exception logic and responsible intervention rather than one-time technology deployment (Chingezi et al., 2026a; Chukwuemeka and Mupa, 2026a).

Relationship to companion research. This paper is one of two linked studies that both use a 70,000-record analytical extract of the DataCo Smart Supply Chain dataset. The companion study, "A Practical Digital-Transformation Maturity and Delivery-Readiness Framework for U.S. SMEs and Small Government Vendors," uses the same public extract only to calibrate the external distribution of a delivery-performance score for a seven-domain organisational maturity instrument tested on a separate synthetic panel. The present paper instead analyses the extract directly at transaction level to benchmark delivery, margin and workflow outcomes and to develop the leakage-controlled readiness model and ERP Operational Capability Index reported below. Order-level descriptive statistics computed from the shared source (for example, the late-delivery rate) are reported consistently across both papers; neither paper's predictive, index or synthetic-panel results depend on the other, and each addresses a distinct measurement question.

1.1 Research questions

1. Which ERP-visible transaction measures provide a defensible operational baseline for delivery, margin, workflow and data-quality performance?
2. How strongly are observable configuration and transaction characteristics associated with late-delivery risk in the public supply-chain extract?
3. Can an out-of-sample delivery-readiness score meaningfully stratify operational outcomes without using realised delivery information as a predictor?
4. How should SMEs and small government vendors evaluate the causal effect of an ERP-enabled operating-model intervention through a prospective pre/post or matched-case design?
5. What implementation controls are required to convert diagnostic analytics into sustained operational capability?

1.2 Contributions

- A construct-to-measure framework that connects ERP implementation practices with operational and vendor-delivery outcomes.

- A transparent empirical benchmark based on 70,000 transaction lines and 44,026 unique orders, with privacy-sensitive data handling.
- A calibrated, leakage-controlled risk model and readiness stratification for late-delivery risk.
- A six-component ERP Operational Capability Index with explicit weights and sensitivity analysis.
- A prospective matched difference-in-differences protocol, KPI dictionary, power scenarios and implementation playbook suitable for SMEs and government vendors.

II. LITERATURE REVIEW AND THEORETICAL FOUNDATION

2.1 ERP value as a complementary organisational capability

ERP systems integrate information and transaction processing across finance, procurement, inventory, sales, production, projects and service operations. Yet integration is not a purely technical property. It requires common definitions, coordinated roles, standardised process hand-offs and disciplined use of shared data. The resource-based view explains why the same software can produce different outcomes across firms: the system becomes valuable when combined with organisational routines, managerial competence, process knowledge and learning mechanisms that are difficult to imitate (Bharadwaj, 2000; Uwizeyemungu and Raymond, 2012).

Early large-sample studies found that ERP investment can be associated with productivity and performance gains, but benefits may emerge over time and vary by the organisation's ability to absorb and exploit the system (Hitt, Wu and Zhou, 2002; Nicolaou, 2004). McAfee (2002) provided longitudinal evidence that enterprise technology adoption was followed by operational improvement, while Hendricks, Singhal and Stratman (2007) reported heterogeneous long-run effects across ERP, supply-chain and customer-relationship systems. These findings reject a simple purchase-to-performance chain, and firm-level evidence on operational-performance outcomes specifically reinforces that the effect is conditional on implementation quality rather than automatic (Madapusi and D'Souza, 2012). The relevant causal sequence is acquisition, implementation, routinised use, adaptation and benefits realisation.

The DeLone and McLean (2003) information-systems success model similarly distinguishes system quality, information quality, service quality, use, satisfaction and net benefits, and subsequent synthesis work has continued to refine the independent variables that best predict net benefits across contexts (Petter, DeLone and McLean, 2013). ERP research extends that logic by showing that system use must reach a level of assimilation and process integration sufficient to affect operational outcomes (Seddon, Calvert and Yang, 2010; Ruivo, Oliveira and Neto, 2014). For this paper, operational capability is defined as the repeatable ability to execute cross-functional work through governed ERP processes while meeting service, control, data-quality and adoption thresholds.

2.2 Process redesign and organisational fit

Process redesign is the mechanism that converts software functionality into improved flow. Configuring an ERP system around unreformed tasks can preserve non-value-adding approvals, duplicate entry and local workarounds. Conversely, imposing a standard template without understanding legitimate operational differentiation can create misfit, shadow systems and resistance. Strong and Volkoff (2010) conceptualise organisation-enterprise-system fit as multidimensional, while Gattiker and Goodhue (2005) show that post-implementation outcomes depend on interdependence and differentiation across organisational units.

The operational literature provides concrete evidence that process and information integration can affect lead time. Cotteleer and Bendoly (2006) observed order lead-time improvement following enterprise technology implementation, and McAfee (2002) documented changes in operational measures over the adoption period. Seethamraju (2013) links ERP use with process agility, while Trkman (2010) identifies strategic alignment, governance, methods, people and culture as critical conditions for business-process-management success. The implication is that cycle time should not be treated as a generic software KPI. It should be decomposed into planned lead time, actual lead time, queue time, hand-off delay, rework and exception ageing.

Implementation research consistently places business-process reengineering among critical success factors (Nah, Lau and Kuang, 2001; Umble, Haft and Umble, 2003; Ram, Corkindale and Wu, 2013; Bradford and Florin, 2003). A rigorous evaluation must therefore document the future-state process and its control logic before interpreting post-go-live metrics. Otherwise, a fall in cycle time may reflect demand mix, shipping policy or staffing changes rather than ERP-enabled redesign.

2.3 Data governance, master data and information quality

Integrated processes are only as reliable as the data that drive them. Wang and Strong (1996) distinguish intrinsic, contextual, representational and accessibility dimensions of data quality. In ERP settings, poor customer, supplier, item, chart-of-account, unit-of-measure and pricing data propagate across modules and create operational errors that appear downstream as rejected invoices, stock discrepancies, delayed fulfilment or unreliable reporting. Xu et al. (2002) show that data-quality problems are central to ERP implementation and require managerial as well as technical responses.

Data governance specifies ownership, standards, approval rights, lineage, retention and issue resolution. It must be operationalised through master-data stewards, validation rules, duplicate controls, mandatory fields, change logs and periodic quality review. The present study therefore treats key-field completeness and consistency as minimum conditions rather than proof of data quality. Completeness can be perfect while values remain inaccurate, stale or semantically inconsistent.

Recent applied research in ERP-enabled supply chains has demonstrated how transaction data can support continuous auditing of delivery variance, margin exceptions, high discounting, workflow anomalies and fraud labels (Chingezi et al., 2026a). That work establishes the value of ERP-visible exception signals; the present article extends the logic from control monitoring to a broader operational-capability framework integrating service design, predictive readiness, implementation governance and benefits evaluation.

2.4 Internal controls and continuous assurance

ERP systems can strengthen controls through configured approval workflows, segregation of duties, role-based access, three-way matching, tolerance limits, audit logs and automated exception reporting. They can also concentrate risk when access is excessive, rules are poorly configured or users circumvent the system. Control effectiveness therefore depends on design, configuration, evidence and remediation rather than on the presence of a control feature, consistent with evidence that ERP's effect on management control systems is mediated by how the system is configured and governed rather than by its adoption alone (Kallunki, Laitinen and Silvola, 2011).

Continuous auditing research moves assurance closer to the transaction. Alles, Kogan and Vasarhelyi (2006) demonstrated continuous monitoring of business-process controls in a large enterprise environment, while Kuhn and Sutton (2010) described the opportunities and challenges of continuous auditing in ERP systems. In operationally intensive settings, the control cycle should include detection, triage, ownership, root-cause analysis, corrective action, closure evidence and recurrence monitoring. This closed-loop logic is consistent with recent data-driven internal-audit frameworks that emphasise repeat-finding prevention and accountable remediation (Chingezi et al., 2026a; Chingezi et al., 2026b).

For small government vendors, controls have a direct delivery dimension. Documentation completeness, schedule and cost visibility, invoice accuracy, subcontractor oversight and corrective-action closure influence both cash conversion and contract performance. The measurement framework in this paper therefore includes delivery reliability and workflow exceptions alongside traditional financial controls.

2.5 Training, adoption and human factors

ERP training is frequently compressed into pre-go-live instruction on transactions and screens. That approach underestimates the cognitive and behavioural demands of integrated work. Users must understand upstream and downstream consequences, data standards, exception handling, approval responsibilities and the reasons behind process changes. Amoako-Gyampah and Salam (2004) show that training, shared beliefs and perceived usefulness matter in ERP environments, while Venkatesh et al. (2003) identify performance expectancy, effort expectancy, social influence and facilitating conditions as determinants of technology use.

Adoption should be measured through role-specific proficiency, transaction completion, error and rework patterns, exception escalation, help-desk demand, workaround use and sustained active use. Sykes, Venkatesh and Johnson (2014) show that advice networks affect employee performance during enterprise-system implementation, highlighting the importance of super users and social support.

Aladwani (2001) similarly positions communication and change management as central to implementation success.

Human-factors research offers a useful cross-domain warning: advanced technology does not remove the need for trained judgement, clear procedures and non-punitive reporting. Recent work by Chukwuemeka and Mupa (2026a) demonstrates how predictive training analytics and human-centred governance can identify learning gaps without allowing automation to substitute for accountable oversight. Applied to ERP, this means that adoption analytics should trigger coaching, process clarification and system improvement rather than punitive monitoring.

2.6 Implementation governance and benefits realisation

Implementation governance aligns scope, decision rights, risk ownership, resources, quality gates and benefits. Project governance is not equivalent to project administration. It requires an accountable sponsor, process owners, data owners, control owners, change authority, acceptance criteria and a transparent mechanism for resolving design trade-offs. Aloini, Dulmin and Mininno (2007) show that ERP implementation risks span organisational, technical, strategic and project dimensions.

Go-live is a transition event, not a benefit. Benefits realisation requires named benefit owners, baseline values, measurement frequency, attribution logic and post-implementation review. Shang and Seddon (2002) classify enterprise-system benefits across operational, managerial, strategic, IT-infrastructure and organisational dimensions, and Chou and Chang (2008) similarly show that implementation factors such as top-management support and process fit shape the extent of realised benefits. Seddon, Calvert and Yang (2010) further show that organisational benefits arise through multiple projects and complementary factors. The current framework consequently separates delivery acceptance from capability stabilisation and sustained benefits.

Post-implementation changes also matter. Nicolaou and Bhattacharya (2006) find that the nature and timing of post-implementation changes influence performance, while Ruivo, Oliveira and Neto (2014) show that SME value depends on stages of use and post-adoption capability. A capability measurement system must therefore continue after stabilisation, incorporating drift, new process variants, control overrides, training turnover and changing business requirements.

2.7 Research gap

The literature provides substantial evidence on ERP success factors and organisational performance, but practical evaluations often remain fragmented. Project teams track budget and go-live; operations teams track service; finance tracks margin and close; internal audit tracks exceptions; and human resources tracks training attendance. The resulting evidence does not show whether the organisation has acquired an integrated capability. This paper addresses the gap by combining these domains into one measurement architecture and by distinguishing three levels of inference: descriptive operational benchmarking, predictive readiness and causal field evaluation.

III. CONCEPTUAL FRAMEWORK AND TESTABLE PROPOSITIONS

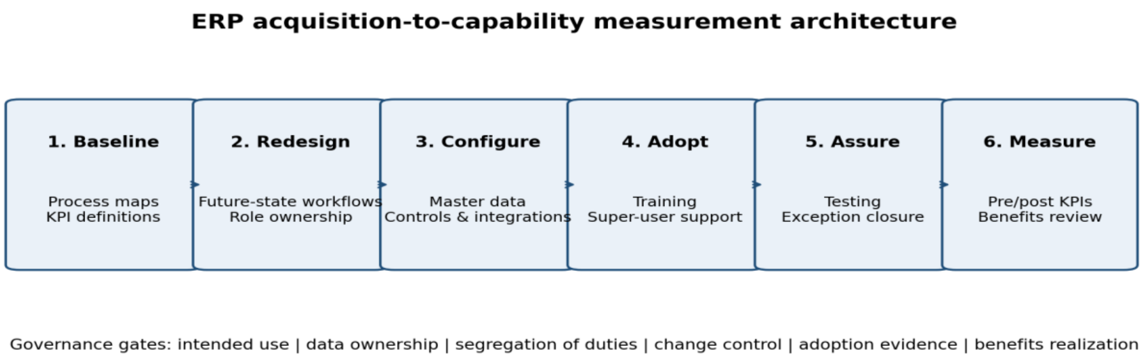

Figure 1. ERP acquisition-to-capability measurement architecture.

The architecture places governance gates across baseline, redesign, configuration, adoption, assurance and measurement. It is vendor-neutral and intended for modular application.

The framework treats ERP capability as a socio-technical production system. Process design determines the intended flow; data governance determines the reliability of shared information; controls constrain and evidence authorised execution; training and support determine competent use; implementation governance coordinates decisions and risks; and benefits measurement determines whether the new operating model produces sustained value. Operational performance is the observable result of the combined system, subject to market, demand, product, logistics and organisational context.

Table 1. Construct-to-measure translation

Capability domain	Operational mechanism	Illustrative measures
Process integration and redesign	Standard work, reduced hand-offs, realistic service policies, clear exception routes	Order-to-ship cycle time; schedule variance; rework; backlog; on-time delivery
Data governance	Owned master data, validation, lineage, consistent definitions and issue closure	Completeness; accuracy sampling; duplicate rate; stale records; reconciliation breaks; reporting latency
Configured controls	Preventive and detective rules embedded in workflows	Approval overrides; segregation-of-duties conflicts; discount exceptions; payment holds; rejected invoices; exception ageing
Training and adoption	Role competence, sustained use and appropriate escalation	Proficiency; active use; error rate; workaround use; support demand; super-user resolution
Implementation governance	Decision rights, scope control, testing, risk ownership and acceptance evidence	Defect closure; change volatility; test coverage; milestone variance; unresolved critical risks
Benefits realisation	Named owners, baselines, attribution and review	Cost-to-serve; working capital; margin; delivery reliability; reporting speed; contract KPI attainment

Note: Measures must be defined consistently before baseline collection. A single metric may be influenced by several domains; causal interpretation requires the prospective design in Section 8.

3.1 Benchmark hypotheses

The public extract permits tests of observable transaction relationships, not tests of training or implementation governance. The following hypotheses are therefore restricted to the benchmark layer:

6. H1: Pre-shipment configuration and order context will discriminate late-delivery risk above chance in a holdout sample.
7. H2: Delivery reliability will differ materially across shipping modes after adjustment for order, market, customer, payment, department and time characteristics.
8. H3: A holdout readiness score based only on pre-outcome information will stratify observed on-time delivery.
9. H4: A composite operational-capability index will produce stable relative rankings under reasonable alternative weights.

3.2 Prospective field propositions

10. P1: ERP implementation accompanied by documented process redesign will produce larger cycle-time and delivery improvements than software configuration without redesign.
11. P2: Data governance and embedded controls will reduce workflow exceptions, invoice rejection, reporting latency and recurrence of control failures.
12. P3: Role-based training, super-user support and adoption monitoring will mediate the relationship between system availability and realised operational performance.

13. P4: Implementation governance and explicit benefits ownership will improve the persistence and auditability of post-go-live gains.

IV. DATA AND METHODS

4.1 Research design

The study uses a sequential empirical and design-science approach. The empirical layer analyses ERP-visible supply-chain transactions to establish operational benchmarks, identify concentration of risk and test a leakage-controlled late-delivery model. The design-science layer translates those findings and the ERP implementation literature into a reusable evaluation protocol, capability index, dashboard logic and implementation playbook.

This design is appropriate because the public dataset contains transaction outcomes but no firm identifier, ERP implementation date, treatment assignment, training measures or governance assessments. It would therefore be methodologically incorrect to label the analysis as an ERP pre/post study. The empirical findings answer what can be measured from integrated transaction records; the prospective protocol specifies how a future study should answer what the implementation caused.

4.2 Dataset and analytical sample

The source is the DataCo Smart Supply Chain for Big Data Analysis dataset, originally released through Mendeley Data and widely mirrored on Kaggle (Constante, Silva and Pereira, 2019). The complete source contains 180,519 records and 53 variables covering provisioning, production, sales and commercial distribution. This study used a publicly available 70,000-record analytical extract. Because the extraction mechanism of the mirror was not documented as probabilistic sampling, results are reported as properties of the extract rather than population estimates for the complete dataset.

Table 2. Data source and analytical sample

Layer	Description	Scale	Use/status
Original dataset	DataCo Smart Supply Chain for Big Data Analysis	180,519 records; 53 variables	Mendeley Data, CC BY 4.0; Kaggle mirrors
Analytical extract	Public 70,000-record extract	70,000 order lines	Used in this study
Order-level sample	Aggregated by unique Order Id	44,026 orders	Delivery, workflow and readiness analysis
Date coverage	Order date	1 January 2015-31 January 2018	Parsed from transaction timestamps
Core domains	Orders, products, markets, shipping, status, discount, sales and profit	Mixed numeric, categorical and date fields	Personally identifying customer fields excluded

4.3 Privacy and data minimisation

Customer names, email addresses, passwords, street addresses, precise coordinates, customer cities and image URLs were removed before analysis. Results are aggregated; no customer-level output is reproduced. The public dataset contains masked or synthetic-looking identity fields, but public availability does not remove the obligation to minimise unnecessary personal data. The analysis retained only operational, commercial, geographic-at-market-level and transaction identifiers required for aggregation.

4.4 Unit of analysis and data-quality checks

Delivery and workflow outcomes were analysed at the order level because multiple order lines can share the same shipping decision and status. Order-line analysis was retained for discount-band and item-level margin descriptions. Twelve delivery and order fields were tested for within-order consistency; no Order Id contained conflicting values for actual shipping days, scheduled days, delivery status, late-delivery label, shipping mode, order date, shipping date, market, region, customer segment, payment type or order status. Key analytic fields were complete in the extract. Product Description was entirely missing and Order Zipcode was missing for most records; both were excluded. Completeness is interpreted narrowly and does not prove semantic accuracy.

Table 3. Operational variables and measurement rules

Variable	Operationalisation	Scale	Analytical role
Late-delivery risk	Binary source indicator	1 if the source record labels the order late; 0 otherwise	Primary delivery outcome
On-time delivery	1 - late-delivery risk	Percentage of orders not labelled late	Primary operational KPI
Schedule variance	Actual shipping days - scheduled shipping days	Days; positive values indicate slippage	Cycle-time conformance
Severe delay	Schedule variance >=2 days	Binary	Material service exception
Negative profit	Aggregated order profit <0	Binary	Margin exposure
Hard workflow exception	Order status is CANCELED, SUSPECTED_FRAUD, PAYMENT_REVIEW or ON_HOLD	Binary	Financial/operational exception proxy
Key-data completeness	All selected operational fields are non-missing	Binary	Minimum data-quality condition
Delivery-readiness score	100 x (1 - predicted late-delivery probability)	0-100	Out-of-sample predictive stratification
EOCI	Weighted composite of six observed KPI rates	0-100	Descriptive capability benchmark

Note: Invoice rejection, reporting latency, training proficiency and governance maturity are not available in the public dataset. They are included in the prospective protocol rather than approximated as observed facts.

4.5 Descriptive, inferential and predictive analysis

Descriptive analysis reports frequencies, means and rates at order and order-line levels. Bootstrap 95% confidence intervals were calculated for central portfolio KPIs using 800 resamples. Cross-segment heat maps display late-delivery, negative-profit and composite capability patterns. Spearman correlation and bias-corrected Cramer's V quantify selected bivariate relationships.

The primary inferential model is a robust logistic regression for order-level late-delivery risk. Predictors were selected to be known at or before dispatch: shipping mode, average discount rate, order value, quantity, line-item count, market, customer segment, payment type, dominant department, year and month. Actual shipping days, delivery status, schedule variance and other realised outcomes were excluded to prevent target leakage. Standard Class, USCA, Consumer, DEBIT and Fan Shop were reference categories. Heteroskedasticity-robust standard errors were used.

A separate regularised logistic model was trained on 70% of orders and evaluated on a stratified 30% holdout set. Numeric predictors were median-imputed and standardised; categorical predictors were most-frequent imputed and one-hot encoded. Performance was assessed using area under the receiver operating characteristic curve (AUC), Brier score, threshold-based classification statistics, calibration intercept and calibration slope. The model was not optimised to claim state-of-the-art prediction; its purpose was to test whether ERP-captured pre-outcome information provides useful operational stratification.

$$logit[P(Late\_i = 1)] = beta\_0 + beta\_1\ ShippingMode\_i + beta\_2\ OrderContext\_i + beta\_3\ Market\_i + beta\_4\ Time\_i$$

4.6 ERP Operational Capability Index

The EOCI is a descriptive group-level index, not a psychometric maturity instrument. It summarises observable operational performance across Market x Shipping Mode cells. The primary weighting allocates 30% to on-time delivery, 20% to schedule adherence, 20% to positive profit, 15% to workflow cleanliness, 10% to low-discount execution and 5% to key-data completeness. The weights emphasise delivery and control outcomes while retaining data quality. An alternative weighting increases schedule-adherence and data-completeness weights; Spearman rank correlation tests sensitivity.

EOCI = 100 x [0.30(OT) + 0.20(SA) + 0.20(PP) + 0.15(WC) + 0.10(LD) + 0.05(DC)]

Table 4. EOCI primary and sensitivity weights

Component	Primary	Alternative	Rationale
On-time delivery (OT)	0.30	0.25	Core customer/contract outcome
Schedule adherence (SA)	0.20	0.25	Process conformance
Positive profit (PP)	0.20	0.15	Commercial sustainability
Workflow clean (WC)	0.15	0.15	Control/exception condition
Low discount (LD)	0.10	0.10	Pricing discipline proxy
Data complete (DC)	0.05	0.10	Minimum information-quality condition

Note: The index supports operational prioritisation. It should not be interpreted as a causal ERP effect or a substitute for validated survey measurement.

4.7 Prospective causal-evaluation design

A future field study should recruit ERP implementations before baseline collection and compare firms or operational units receiving the full capability intervention with matched comparators. Matching variables should include sector, employment, transaction volume, product/service complexity, baseline digital maturity, baseline outcome levels and pre-intervention trends; propensity-score matching or weighting on these variables, following current best-practice guidance, is preferred to ad hoc nearest-neighbour selection (Austin and Stuart, 2015). Where implementation timing varies, the analysis should use modern staggered-adoption estimators rather than a single two-way fixed-effects coefficient (Callaway and Sant'Anna, 2021).

Y_it = alpha + beta_1 Treat_i + beta_2 Post_t + beta_3(Treat_i x Post_t) + gamma X_it + mu_i + tau_t + epsilon_it

The interaction coefficient beta_3 estimates the average change in the treated group beyond contemporaneous change in the comparison group, subject to the parallel-trends assumption. Standard errors should be clustered at the firm or implementation-unit level. Event-study coefficients should test pre-trends and reveal whether benefits emerge immediately, during stabilisation or only after adoption matures (Bertrand, Duflo and Mullainathan, 2004).

V. RESULTS

5.1 Analytical sample and baseline capability profile

The extract contained 70,000 order lines aggregated to 44,026 unique orders. Mean actual shipping time was 3.50 days against a mean scheduled time of 2.93 days, producing mean slippage of 0.56 days and a median variance of 1 day. On-time delivery was 45.3% (95% bootstrap CI 44.9%-45.8%). Severe delay affected 23.6% of orders. Negative-profit orders represented 19.6% (95% CI 19.2%-20.0%), while hard workflow exceptions affected 10.8% (95% CI 10.5%-11.1%).

Table 5. Portfolio baseline KPIs

KPI	Result	Interpretation
Order lines	70,000	Transaction detail
Unique orders	44,026	Delivery/workflow unit
On-time delivery	45.3%	95% CI 44.9%-45.8%
Late delivery	54.7%	Primary adverse outcome

KPI	Result	Interpretation
Severe delay (>=2 days)	23.6%	Material slippage
Negative-profit orders	19.6%	95% CI 19.2%-20.0%
Hard workflow exception	10.8%	95% CI 10.5%-11.1%
Mean actual vs scheduled	3.50 vs 2.93 days	Mean variance 0.56 days
Key analytic data complete	100.0%	Selected fields only

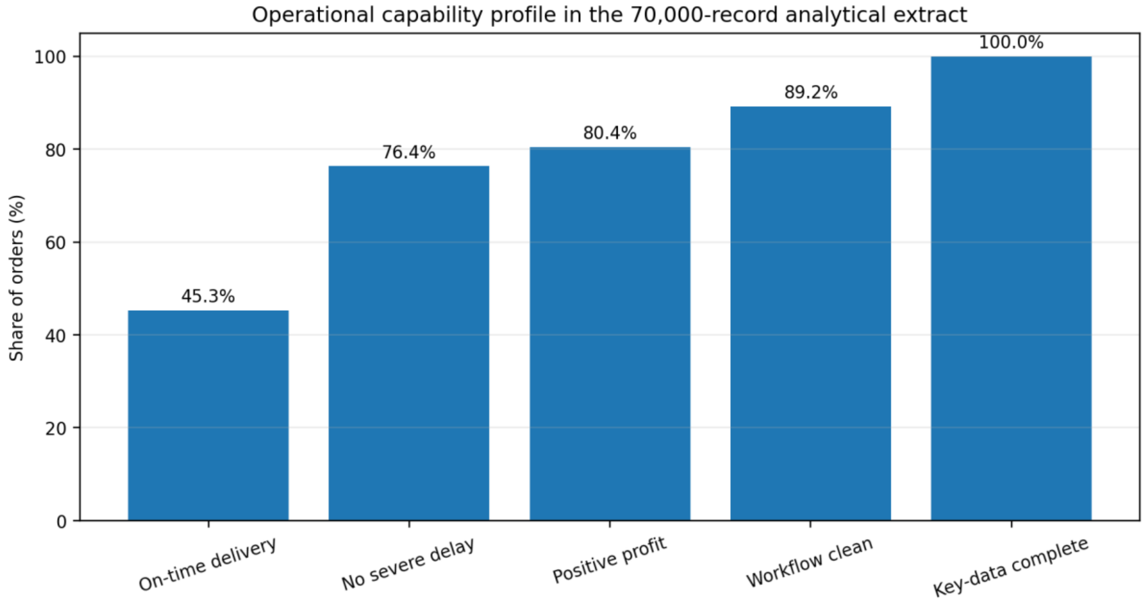

Figure 2. Operational capability profile in the analytical extract.
Data completeness is limited to selected analytic fields and should not be interpreted as proof of master-data accuracy.

5.2 Shipping configuration and delivery performance
Shipping configuration produced the clearest performance separation. First Class orders were late in 95.3% of cases, Second Class in 76.6%, Same Day in 46.0% and Standard Class in 37.9%. The pattern is operationally counterintuitive only if service labels are mistaken for capability. In the extract, First Class was scheduled for one day but realised two days on average; Second Class was scheduled for two days but realised 3.99 days. Standard Class was scheduled for four days and realised 3.99 days. Thus, faster promised classes were structurally misaligned with observed fulfilment capacity.

Table 6. Delivery performance by shipping mode

Mode	Orders	Late	Severe delay	Actual days	Scheduled	Variance
First Class	6,728	95.3%	0.0%	2.00	1.00	1.00
Second Class	8,588	76.6%	59.8%	3.99	2.00	1.99
Same Day	2,400	46.0%	0.0%	0.48	0.00	0.48
Standard Class	26,310	37.9%	20.0%	3.99	4.00	-0.01

Note: Severe delay is schedule variance of at least two days. Shipping modes are configuration categories in the source data, not experimental treatments.

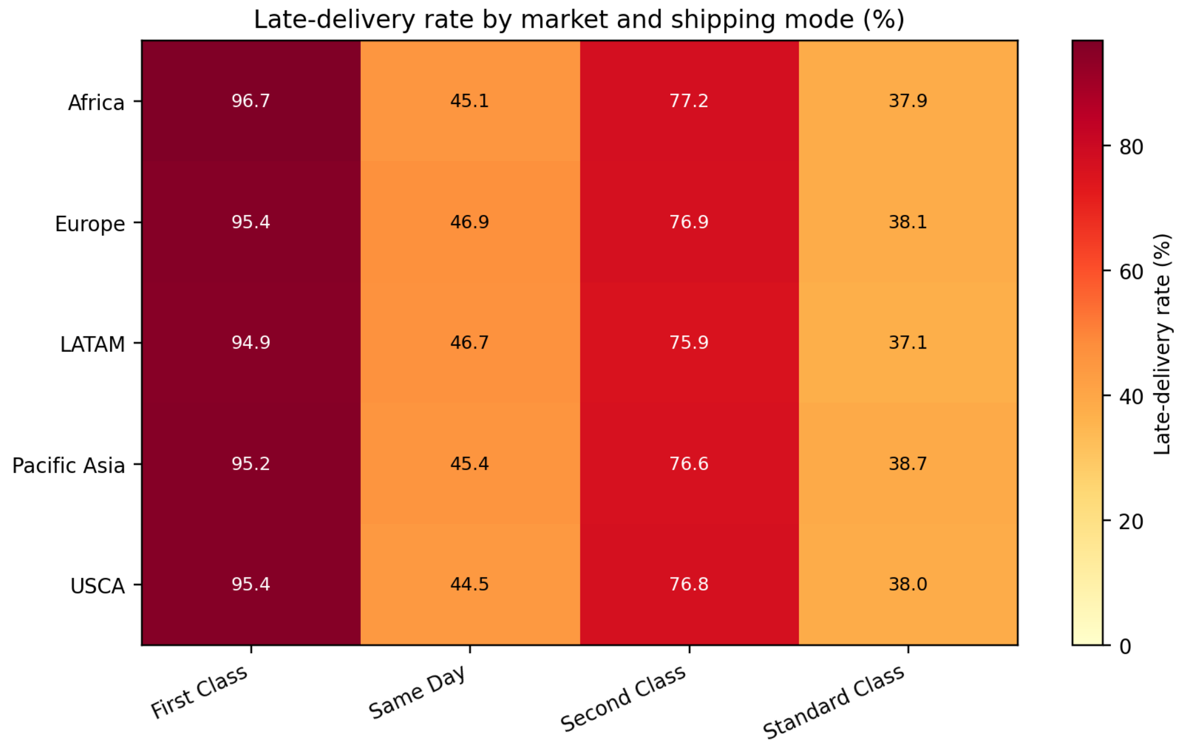

Figure 3. Late-delivery rate by market and shipping mode.

The shipping-mode gradient is consistent across markets, while broad market differences are comparatively small.

Table 7. Selected adjusted late-delivery associations

Predictor	Adjusted OR	95% robust CI	p-value
First Class vs Standard	34.22	30.56-38.32	<0.001
Second Class vs Standard	5.42	5.12-5.73	<0.001
Same Day vs Standard	1.40	1.28-1.52	<0.001
Discount rate (+5 percentage points)	1.01	0.99-1.03	0.226
Log order value (+1)	1.02	0.98-1.06	0.378
Quantity (+1)	0.99	0.98-1.01	0.238
Line-item count (+1)	1.02	0.98-1.07	0.281

Note: Robust logistic regression adjusted for market, customer segment, payment type, department, year, month, discount, order value, quantity and line-item count. Reference shipping mode: Standard Class.

Adjusted results reinforced the descriptive finding. Relative to Standard Class, the odds of late delivery were 34.22 times higher for First Class, 5.42 times higher for Second Class and 1.40 times higher for Same Day. Discount, order value, quantity and line-item count were not statistically important in the selected model once configuration and context were included. Cramer's V for shipping mode and late delivery was 0.457, indicating a substantial association for a categorical operational variable.

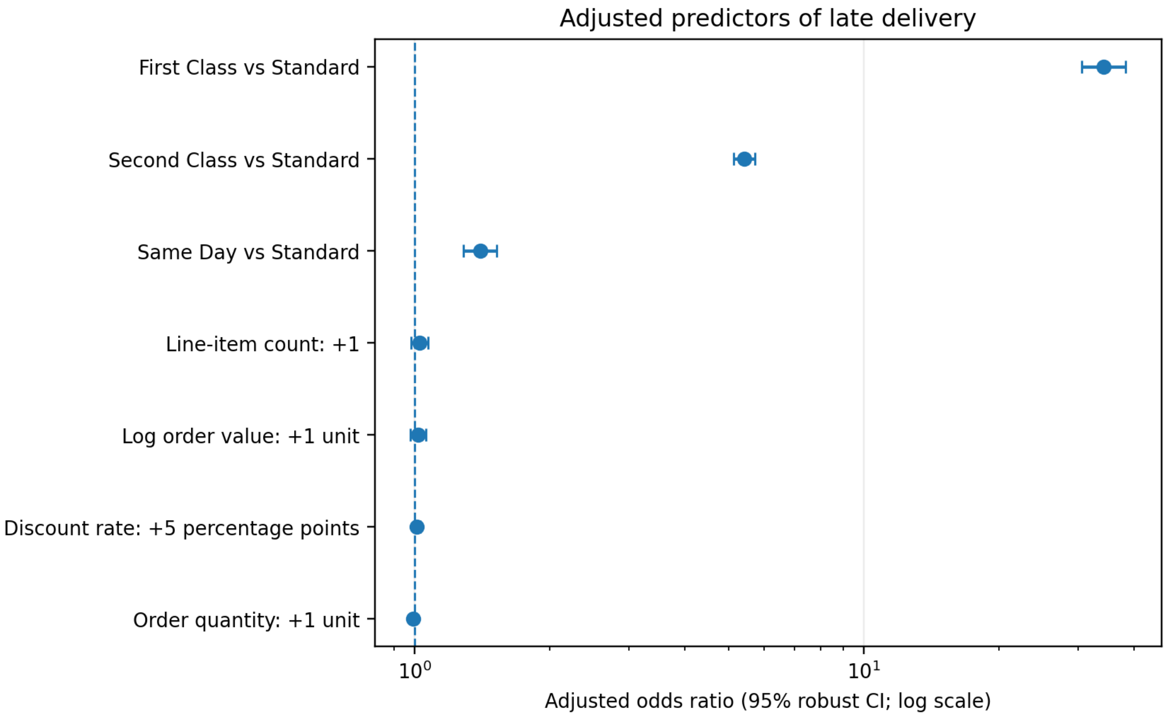

Figure 4. Adjusted predictors of late delivery.
Odds ratios are shown on a logarithmic scale. Confidence intervals use heteroskedasticity-robust standard errors.

5.3 Market variation
Late-delivery rates ranged narrowly from 53.9% in LATAM to 55.1% in Pacific Asia and Europe. Negative-profit rates ranged from 18.8% to 20.5%, and hard workflow exceptions ranged from 10.2% to 11.1%. These small broad-market differences contrast with the large shipping-mode differences. The result suggests that an ERP performance dashboard should avoid escalating geography merely because aggregate volume is high; it should isolate the process configuration or service promise that explains risk.

Table 8. Operational KPIs by market

Market	Orders	Late	Severe delay	Negative profit	Hard exception	Avg. order
Pacific Asia	10,717	55.1%	24.3%	18.8%	10.6%	$271.86
Europe	12,229	55.1%	23.6%	19.8%	10.7%	$305.03
USCA	6,130	54.7%	23.7%	20.5%	11.1%	$290.25
Africa	2,759	54.2%	23.5%	20.1%	10.2%	$291.74
LATAM	12,191	53.9%	23.2%	19.5%	11.0%	$294.40

5.4 Discounting and margin exposure
At order-line level, the Spearman correlation between discount rate and profit was -0.053. Negative-profit rates remained close to 18% across discount bands, while mean profit declined from $23.93 in the 0-5% band to $18.10 above 20%. In the adjusted order-level negative-profit model, a five-percentage-point discount increase had an odds ratio of 1.003 (p=0.758). The extract therefore does not support a simple claim that discount rate alone explains negative-profit incidence. Margin governance requires product cost, freight, returns, allocation and price-authorisation context that is not fully represented here.

Table 9. Margin outcomes by discount band

Discount band	Lines	Negative profit	Mean profit	Mean profit ratio	Mean discount
0-5%	19,476	18.7%	$23.93	0.122	2.0%
5-10%	15,540	18.1%	$23.23	0.124	6.8%
10-20%	27,173	18.4%	$21.69	0.124	14.4%
>20%	7,811	18.7%	$18.10	0.122	22.5%

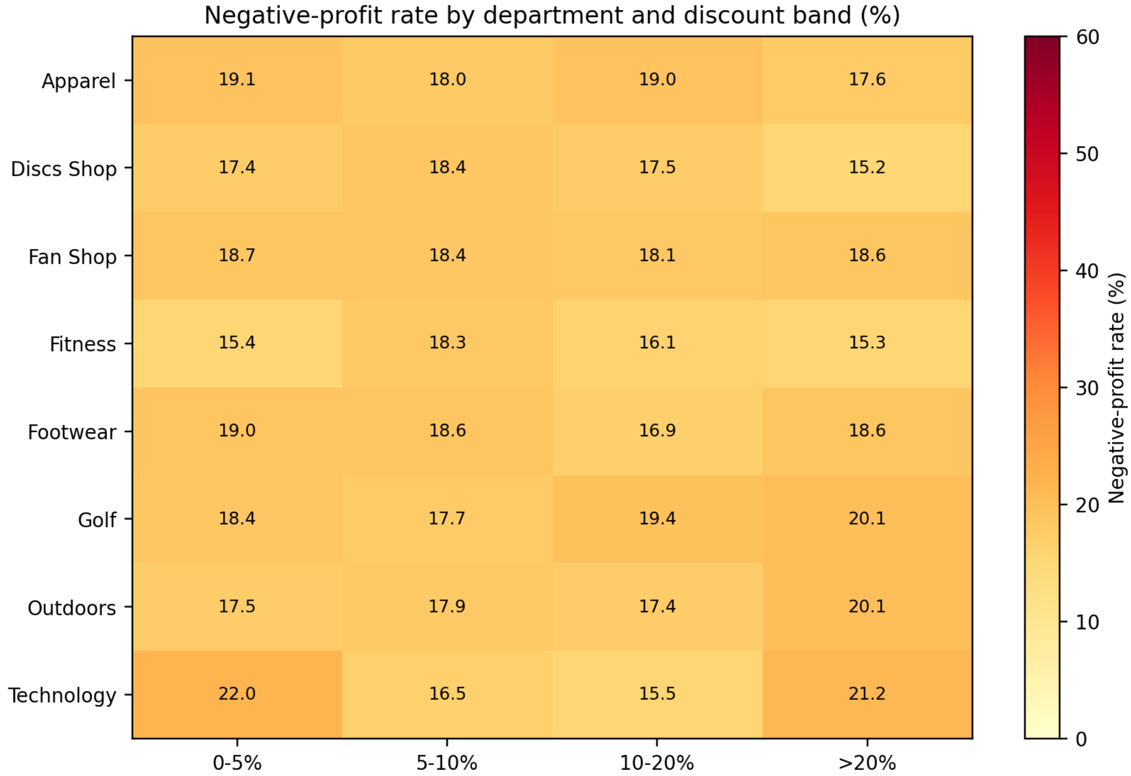

Figure 5. Negative-profit rate by department and discount band.
Department-level variation is modest and non-monotonic. The heat map supports targeted investigation, not causal attribution to discounting.

5.5 Predictive readiness model
The holdout set contained 13,208 orders with a late-delivery prevalence of 54.7%. The model achieved AUC=0.731 and Brier score=0.197. At a 0.50 threshold, accuracy was 69.2%, precision 82.9%, recall 55.1% and F1=0.662. Calibration was close to ideal in aggregate, with intercept 0.008 and slope 0.967.

Table 10. Holdout model performance

Metric	Value	Interpretation
Training orders	30,818	70% stratified split
Holdout orders	13,208	30% stratified split
AUC	0.731	Discrimination
Brier score	0.197	Probability accuracy; lower is better
Calibration intercept	0.008	Ideal=0

Metric	Value	Interpretation
Calibration slope	0.967	Ideal=1
Accuracy	69.2%	Threshold=0.50
Precision	82.9%	Positive predictive value
Recall	55.1%	Sensitivity
F1 score	0.662	Precision-recall balance

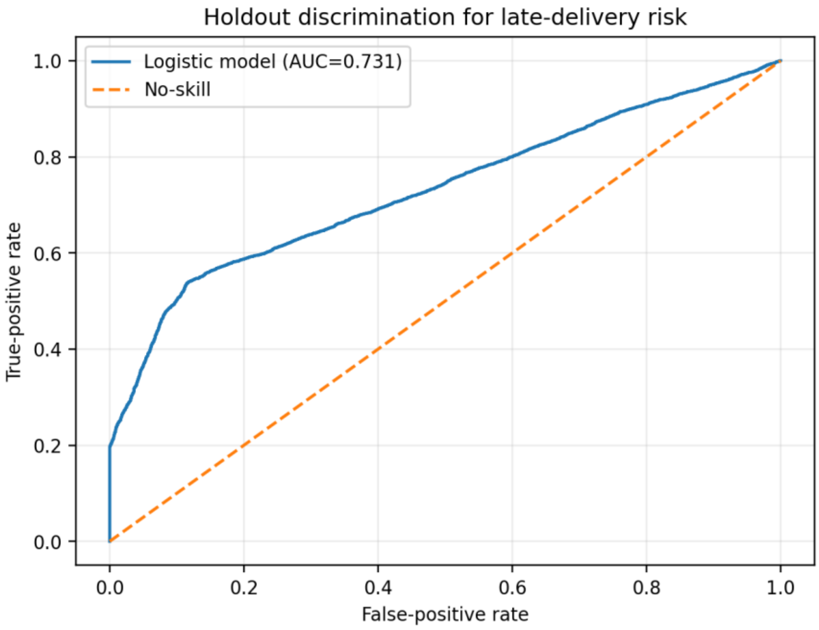

Figure 6. Holdout discrimination for late-delivery risk.

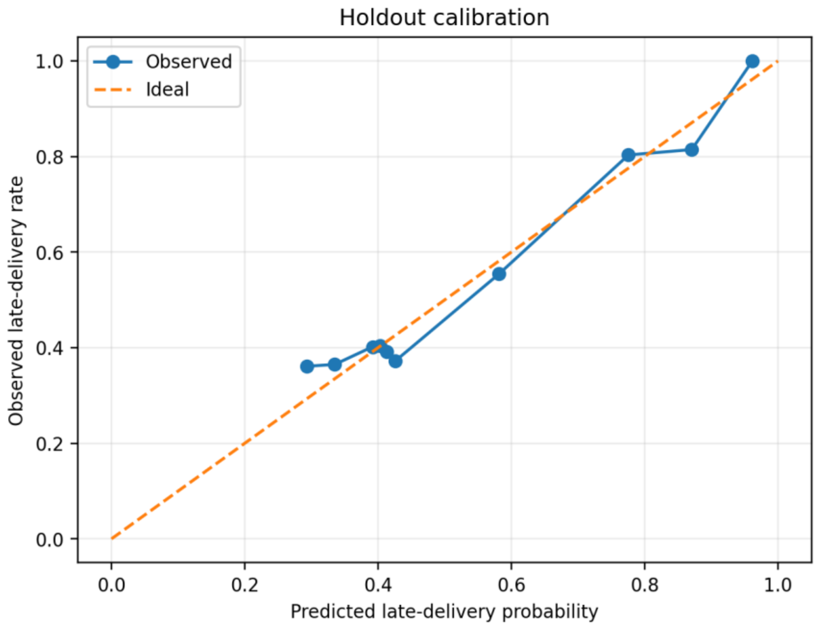

Figure 7. Holdout calibration of late-delivery probabilities.

Calibration was evaluated in quantile bins. Site-specific recalibration would be required before operational use in another organisation.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722515

5.6 Readiness quartiles

Orders were assigned a delivery-readiness score equal to 100 minus their out-of-sample predicted late-delivery probability. Observed on-time delivery rose from 11.2% in the lowest quartile to 62.9% in the highest. The Q4-Q1 difference was 51.6 percentage points (95% bootstrap CI 49.5-53.6). This is a stratification result, not a treatment effect: the score identifies configurations and contexts associated with risk, but changing a score component will not necessarily produce the full observed difference.

Table 11. Holdout outcomes by delivery-readiness quartile

Quartile	Orders	Mean score	On-time	Mean variance	Severe delay	Negative profit
Q1 lowest	3,302	11.0	11.2%	1.39	23.7%	18.3%
Q2	3,302	44.4	47.0%	0.83	32.5%	20.4%
Q3	3,302	59.5	60.2%	0.03	20.0%	19.3%
Q4 highest	3,302	67.1	62.9%	0.07	20.2%	19.6%

Note: Readiness score is a predictive operational score derived without realised delivery fields. It is not a survey-based organisational maturity score.

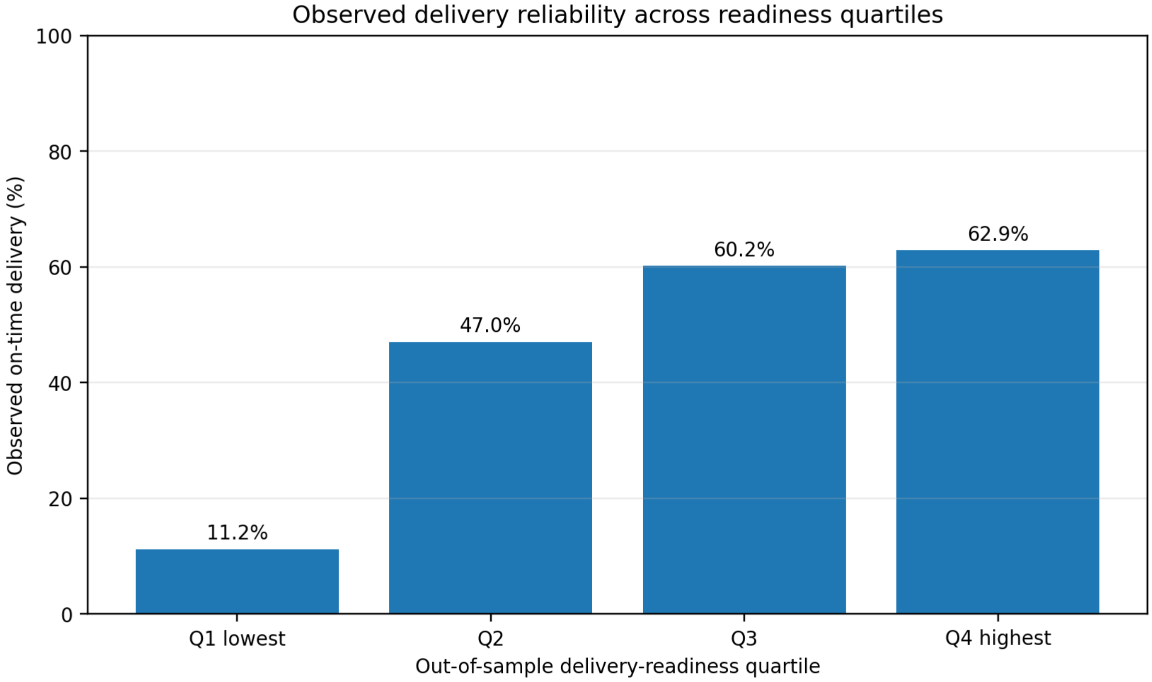

Figure 8. Observed on-time delivery across readiness quartiles.

5.7 ERP Operational Capability Index

EOCI results separated the Market x Shipping Mode cells into a stable operational pattern. Standard Class cells scored approximately 74.1-74.8 and were classified Strong; Same Day scored 69.3-71.1; Second Class scored 54.3-55.2 and was Fragile; First Class scored 44.7-45.5 and was Critical. The near-identical ranking across markets confirms that service configuration, rather than broad geography, drove the composite. Alternative weights produced a Spearman rank correlation of 0.998 with the primary index.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722515

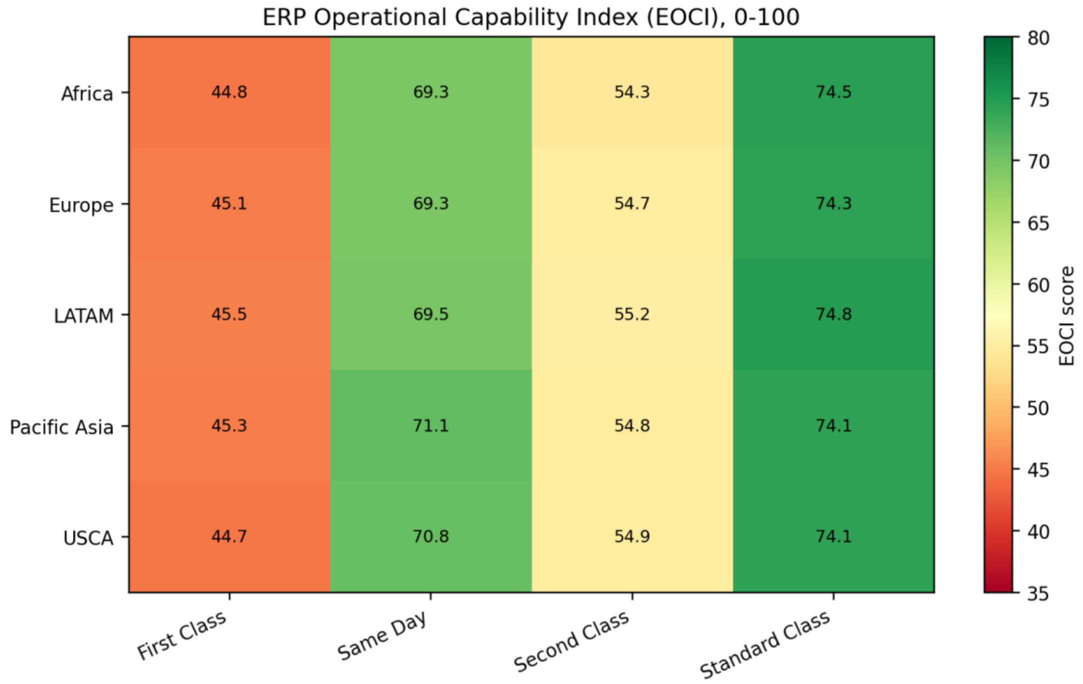

Figure 9. ERP Operational Capability Index by market and shipping mode.
EOCI is a descriptive benchmark incorporating observed outcomes and should not be used as a causal predictor in the same sample.

Table 12. Highest and lowest EOCI cells

Position	Market	Shipping mode	Orders	EOCI	Tier
Highest	LATAM	Standard Class	7,355	74.8	Strong
Highest	Africa	Standard Class	1,669	74.5	Strong
Highest	Europe	Standard Class	7,257	74.3	Strong
Highest	Pacific Asia	Standard Class	6,396	74.1	Strong
Highest	USCA	Standard Class	3,633	74.1	Strong
Lowest	USCA	First Class	921	44.7	Critical
Lowest	Africa	First Class	394	44.8	Critical
Lowest	Europe	First Class	1,909	45.1	Critical
Lowest	Pacific Asia	First Class	1,620	45.3	Critical
Lowest	LATAM	First Class	1,884	45.5	Critical

VI. DISCUSSION

6.1 Acquisition is an input; capability is an operating outcome

The results support the paper’s central distinction. Transaction integration creates visibility, but the visibility reveals whether process promises, workflow rules and operating capacity are aligned. The extract’s First Class category illustrates the point: the system consistently scheduled one day and consistently realised two days, producing a 95% late rate. The problem is not lack of digital recording. It is that the recorded service policy and observed operating capability are misaligned. This is a property of the DataCo extract's own service-level design rather than a general claim about First Class shipping or about ERP-enabled operations elsewhere; the value of the example is illustrative of what transaction-level analytics can surface, not a portable benchmark.

An organisation that treats go-live as success may continue to report such transactions accurately without correcting the underlying promise, capacity or process. Operational capability requires a feedback mechanism that detects the pattern, assigns ownership, tests root causes, adjusts process or capacity and monitors recurrence. That mechanism is complementary to the ERP system; it does not arise automatically from the software.

The finding is consistent with resource-based and post-implementation research. Enterprise systems provide a platform for integration, but benefits depend on managerial, process and organisational capabilities (Uwizeyemungu and Raymond, 2012; Ruivo, Oliveira and Neto, 2014). The practical implication is that implementation assurance should ask not only whether transactions post correctly, but whether the system produces stable, reliable and governable outcomes.

6.2 Process redesign and service-policy realism

The shipping-mode gradient is a process-design result with direct managerial relevance. A service class is a promise embedded in master data and order logic. If the promise is systematically infeasible, the ERP system will automate failure and generate recurring exceptions. Process redesign must therefore include service-level segmentation, cutoff times, inventory availability, pick-pack capacity, carrier hand-off, route constraints and exception rules.

The appropriate response is not automatically to lengthen every promise. Some customers or contracts may require faster service, in which case the operating model must supply the capacity. The key is to align the promise with evidence. Scenario analysis should estimate the cost and operational requirements of each service level, while dashboards should distinguish controllable internal delay from carrier or external delay. The broad-market results show why this analysis should begin with process configuration rather than generic geographic escalation.

6.3 Data governance and control implications

The extract had complete key analytic fields and no within-order conflicts in core shipment attributes, which enabled reliable aggregation. That is a useful minimum condition. It does not establish that product, cost, customer, inventory or pricing master data were accurate. In practice, ERP capability measurement should pair automated completeness and consistency rules with sampled accuracy testing, source reconciliation and ownership of unresolved issues.

Margin results illustrate the danger of simplistic controls. Higher discount bands had lower mean profit, but discount rate alone did not materially explain negative-profit incidence after adjustment. A control that blocks all discounts above a fixed threshold may therefore create friction without isolating the true drivers. Better control design combines discount authority, contribution margin, product cost, freight, customer terms, returns and promotional context. Continuous auditing should route exceptions by materiality and recurrence, not only by a single field threshold.

Workflow exceptions affected approximately one in nine orders. A mature operating model should classify these exceptions, measure ageing, assign accountable owners, capture root cause and confirm closure. Repeated exceptions should trigger process or configuration changes. This closed-loop design is more important than the raw existence of an exception report.

6.4 Training and adoption

The public dataset cannot measure training, but the readiness findings clarify what training should accomplish. Users need to understand which service promises are feasible, how to select the correct workflow, when to escalate exceptions and how their data entry affects downstream delivery, billing and reporting. Training attendance is therefore an input metric. Competence and process adherence are closer to capability.

Adoption analytics should be role-specific and proportionate. A warehouse user may be evaluated on scan completeness, timely confirmation and exception escalation; a project manager on milestone updates and cost-to-complete accuracy; an accounts-receivable user on billing timeliness, rejection resolution and cash-application accuracy. The purpose should be improvement, not surveillance. Chukwuemeka and Mupa’s (2026a) predictive training framework is relevant because it treats analytics as a means of early support and targeted intervention rather than as a substitute for professional judgement.

6.5 Implications for SMEs

SMEs should resist the false choice between a comprehensive transformation programme and no governance. A modular approach can establish a minimum viable operating system: one end-to-end process, a controlled master-data set, a small number of preventive and detective controls, role-based training, a super-user network and a concise KPI dashboard. The design can then expand as evidence and capacity grow.

The most important implementation decision is often scope discipline. An SME that configures too many modules, customisations and reports at once may dilute process ownership and training. A phased design should prioritise the transaction streams that affect cash, customer service and regulatory or contractual obligations. Each phase should have a baseline, acceptance criteria and a benefits owner.

The findings also support fit-for-purpose analytics. The holdout model's AUC of 0.731 is useful for triage but not sufficient for automated decisions. SMEs do not need a complex model for every process. A transparent rule or logistic score may be preferable when it is calibrated, monitored and connected to an actionable workflow. Recent applied research on workflow optimisation similarly cautions that analytical sophistication must remain aligned with process purpose and organisational capacity.

6.6 Implications for small government vendors

For government vendors, operational capability has an evidentiary dimension. A vendor must be able to show what was ordered, approved, delivered, billed, corrected and closed. ERP-enabled performance therefore includes documentation completeness, milestone visibility, cost and schedule variance, invoice acceptance, subcontractor traceability, change control and corrective-action closure.

The proposed framework does not certify compliance or contract eligibility. It provides a measurement structure that can support reliable performance. A vendor-readiness dashboard should distinguish contract requirements from internal targets, preserve source documentation and record reviewer decisions. Where cybersecurity, controlled information or specialised accounting requirements apply, qualified professionals must determine the relevant standards and system boundaries.

The acquisition-capability distinction is especially valuable in public contracting because a software demonstration can create false confidence. Contracting performance depends on daily execution and evidence. The field protocol in Section 8 can be adapted to compare vendor units or cohorts before and after a governed ERP-enabled intervention.

VII. IMPLEMENTATION PLAYBOOK: CONVERTING ERP INTO OPERATIONAL CAPABILITY

The following playbook translates the findings into six implementation stages. It is designed for vendor-neutral use and can be scaled from a single process pilot to a multi-module programme.

Table 13. ERP acquisition-to-capability implementation playbook

Stage	Core activities	Required evidence	Primary failure risk
1. Governed baseline	Define intended use; appoint sponsor, process/data/control owners; map current flow; freeze KPI definitions	Baseline pack, process map, data inventory, risk register, measurement plan	No agreed baseline; ambiguous ownership; metric disputes
2. Future-state redesign	Remove duplication; define standard work, service rules, exception routes and role boundaries	Future-state map, RACI, requirements, service catalogue, control points	Configuring existing dysfunction; over-customisation
3. Data and control design	Clean and govern master data; configure validation, access, approvals, tolerance limits and logging	Data standards, migration evidence, access matrix, control catalogue, reconciliation plan	Poor migration; excessive access; weak evidence

Stage	Core activities	Required evidence	Primary failure risk
4. Build, test and cutover	Configure/integrate; execute unit, process, security, migration and user-acceptance tests; plan rollback	Traceability matrix, test results, defect log, cutover plan, acceptance decision	Low process coverage; unresolved critical defects
5. Adoption and stabilisation	Deliver role-based training; establish super users; monitor proficiency, workarounds, support and exceptions	Competence evidence, adoption dashboard, support trends, issue closure	Attendance mistaken for competence; shadow systems
6. Benefits and lifecycle governance	Compare with baseline; validate attribution; review controls, data quality and drift; assign benefit actions	Benefits report, KPI dashboard, control assurance, lessons, roadmap	Premature success claims; benefits without owners

Note: Each stage should have explicit entry and exit criteria. An organisation may iterate between stages as defects or changing requirements emerge.

7.1 Minimum viable KPI dashboard

Table 14. Minimum viable ERP capability dashboard

Domain	KPI	Definition	Frequency	Accountable owner
Customer/contract	On-time delivery	Orders delivered within defined promise / delivered orders	Daily/weekly	Process owner
Flow	Order-to-ship cycle time	Shipment timestamp - order acceptance timestamp	Daily/weekly	Operations
Conformance	Schedule variance	Actual cycle time - planned cycle time	Daily	Operations
Inventory	Inventory accuracy	1 - absolute count variance / expected quantity	Cycle count/monthly	Inventory owner
Finance	Reporting latency	Report availability - period close/cutoff	Monthly	Controller
Billing	Invoice rejection	Rejected invoices / submitted invoices	Weekly/monthly	Billing owner
Delivery	Milestone attainment	Milestones on time / due milestones	Weekly/monthly	Project/contract manager
Controls	Exception ageing	Median days from detection to closure	Weekly	Control owner
Data	Master-data defects	Validated defects / records tested	Weekly/monthly	Data steward
Adoption	Role proficiency	Users passing task-based assessment / users assessed	At go-live + quarterly	Training/process owner

Domain	KPI	Definition	Frequency	Accountable owner
Adoption	Workaround rate	Transactions completed outside approved process / total	Weekly	Process owner
Benefits	Realised benefit	Verified benefit achieved / approved benefit target	Monthly/quarterly	Benefit owner

Note: Definitions must specify numerator, denominator, exclusions, timestamp source, materiality and treatment of reopened exceptions. Targets should be process- and contract-specific.

7.2 Governance gates

- Intended-use gate: the process, users, decisions and exclusions are documented.
- Data gate: critical master and transactional data meet defined completeness, accuracy and reconciliation thresholds.
- Control gate: role access, approvals, logging and exception ownership are tested.
- Process gate: end-to-end scenarios, including failures and reversals, pass acceptance criteria.
- Adoption gate: users demonstrate role competence and support capacity is operational.
- Benefits gate: baselines, benefit owners and attribution methods are approved before outcome reporting.
- Lifecycle gate: change control, monitoring, periodic access review and KPI recalibration are active.

VIII. PROSPECTIVE MATCHED-CASE FIELD VALIDATION PROTOCOL

8.1 Study population and intervention

The proposed field study should recruit U.S. SMEs and small government vendors planning finance, inventory, procurement, project, service or integrated ERP implementation. The treated condition is not merely software go-live. It is a defined capability package consisting of current-state diagnosis, process redesign, data governance, control configuration, role-based training, implementation governance and post-go-live benefits review. Comparator organisations may use existing systems or implement software without the full package, provided their condition is documented rather than assumed.

Table 15. Prospective field-study specification

Component	Recommended treatment
Design	Matched cohort with difference-in-differences; staggered adoption where necessary
Baseline	Minimum six months of monthly KPIs; preferably twelve months for seasonality and pre-trend assessment
Follow-up	At least twelve months after go-live; report 0-3 month stabilisation separately
Matching	Sector, size, transaction volume, product/service complexity, baseline digital maturity, baseline KPI level, pre-trend
Primary outcomes	Cycle time, inventory accuracy, reporting latency, invoice rejection, on-time delivery and active adoption
Secondary outcomes	Margin, exception recurrence, schedule/cost variance, support demand, documentation completeness, user proficiency
Intervention fidelity	Completion and quality of redesign, data, control, training and governance components

Component	Recommended treatment
Analysis	Event study and DID with firm/unit and time effects; clustered standard errors; sensitivity to matching and attrition
Qualitative layer	Structured interviews and case studies to explain mechanisms, workarounds and implementation context
Transparency	Pre-specified protocol, data dictionary, codebook, exclusion flow and limitations

8.2 Measurement schedule

Table 16. Recommended longitudinal measurement schedule

Period	Phase	Measurement focus
T-6 to T-1	Baseline	Collect monthly KPIs; process maps; maturity survey; data/control tests; pre-trends
T0	Readiness/cutover	Acceptance evidence; critical defects; access; training competence; cutover decision
T+1 to T+3	Stabilisation	Daily/weekly service, defects, workarounds, support, exceptions; avoid premature benefit claims
T+4 to T+6	Assimilation	Monthly adoption, process conformance, data quality, control and operational outcomes
T+7 to T+12	Benefits	Sustained outcomes, benefit validation, control recurrence, staff turnover effects
T+13 onward	Lifecycle	Upgrades, new modules, drift, access recertification, benchmark refresh

8.3 Power scenarios

The table below provides simple two-arm order-level power scenarios for detecting reductions from a 55% late-delivery baseline. The independent-order calculation is optimistic when observations are clustered within firms. The second sample column applies an illustrative design effect of 1.9, chosen only as a round, moderately conservative placeholder and not derived from a pilot ICC or prior clustered study of this population; a final calculation should use pilot estimates of intraclass correlation, cluster count, unequal cluster size, attrition and the DID structure.

Table 17. Illustrative power scenarios for late-delivery reduction

Baseline	Post	Reduction	Orders/arm (iid)	Orders/arm (DE=1.9)
55.0%	50.0%	5 pp	1,565	2,973
55.0%	45.0%	10 pp	392	744
55.0%	40.0%	15 pp	173	329

Note: Two-sided alpha=0.05 and 80% power. These are planning scenarios, not a substitute for cluster-aware study design.

8.4 Case-validation template

Table 18. Standardised case-study reporting template

Section	Required content
Context	Sector, size, operating model, implementation scope, vendors, constraints
Baseline	Process map, KPI history, known data/control issues, digital maturity
Intervention	Specific redesign, data, control, training and governance activities
Fidelity	What was completed, delayed, modified or not adopted
Outcomes	Pre/post trends with comparison; confidence intervals; operational and financial results
Mechanisms	Which process or behavioural changes plausibly produced the outcomes
Counterfactual risks	Demand, staffing, product, policy, seasonality or supplier changes
Limitations	Missing data, measurement changes, short follow-up, selection and spillovers
Transferability	Conditions under which the case is likely or unlikely to generalise

IX. LIMITATIONS AND RESEARCH AGENDA

14. The 70,000-record extract is not documented as a random sample of the complete 180,519-record dataset. The results may differ from the complete source and should be reproduced on the original release.
15. The source is a single historical e-commerce supply-chain dataset. It does not represent all ERP modules, industries, SMEs or government vendors.
16. The dataset contains no ERP implementation date, firm identifier, training exposure, governance assessment or pre-intervention treatment assignment. The paper therefore makes no causal claim that ERP implementation produced the observed outcomes.
17. Late-delivery risk is a source label. Although it is consistent with the schedule-variance pattern, the study cannot independently verify external carrier events or all business rules used to create the label.
18. Order-level profit was aggregated from line values. Cost allocation, freight, returns, rebates and accounting policies are not fully observable, limiting margin interpretation.
19. Key-field completeness was high, but the analysis could not validate semantic accuracy, timeliness or master-data ownership. Missingness in excluded fields shows that completeness depends on the intended use.
20. The delivery-readiness model was validated by a random holdout from the same extract. External, temporal and site-level validation is required before operational deployment. Model monitoring should include calibration, drift, subgroup performance and workflow impact.
21. EOCI includes outcomes and is intended for descriptive benchmarking. It must not be used as an independent causal predictor in the same data. A future maturity instrument should undergo expert review, reliability testing, factor analysis and external validation.
22. The power table uses simplified independent-proportion assumptions and an illustrative design effect. A field trial must account for clustering, repeated measures, implementation heterogeneity and attrition.

The research agenda should therefore prioritise multi-organisation field data, validated measures of process redesign and governance fidelity, direct adoption telemetry, invoice and reporting outcomes, external validation of risk models, and mechanism-oriented case studies. A particularly valuable extension would compare full capability interventions with software-only implementations, allowing the contribution of complementary organisational practices to be estimated directly.

X. CONCLUSION

This study develops a rigorous and practical answer to a common digital-transformation problem: buying an ERP system is not the same as acquiring the capability to deliver through it. The transaction benchmark shows that integrated data can reveal persistent operational misalignment. In the analytical extract, faster shipping labels were associated with much higher late-delivery risk because planned service levels exceeded realised capability. A transparent holdout model provided useful risk stratification and was well calibrated in-sample, while the EOCI produced stable descriptive rankings under alternative weights.

The empirical contribution is intentionally bounded. The data support operational benchmarking and predictive readiness, not causal claims about implementation. The paper's broader contribution is therefore the evaluation architecture: baseline the current operation; redesign processes; govern data and controls; develop user competence; enforce implementation gates; and measure benefits against an explicit counterfactual. This case-based, mechanism-oriented view of why some ERP initiatives succeed over the long run while others stall is consistent with recent qualitative causal-network evidence on enterprise-system outcomes (Zerbino et al., 2021). The matched-case protocol, KPI dictionary and playbook make that architecture testable.

For SMEs and small government vendors, the practical message is direct. ERP value is realised when the organisation can repeatedly execute work, produce reliable evidence, resolve exceptions and improve performance. Software is the platform. Operational capability is the outcome.

Declarations

Funding: No external funding was received for this study.

Conflict of interest: The author declares no conflict of interest.

Ethics: The study used a public secondary dataset and did not involve interaction with human participants. Personally identifying fields were excluded from the analytical dataset and no individual-level results are reported.

Data availability: The original DataCo Smart Supply Chain dataset is available from Mendeley Data under CC BY 4.0 and is mirrored on Kaggle. This manuscript analyses a public 70,000-record extract and reports the extraction limitation explicitly.

Code availability: The analysis workflow used reproducible Python procedures for aggregation, descriptive analysis, logistic regression, bootstrap intervals, predictive validation and visualisation. A publication submission should deposit cleaned, non-identifying code and a data dictionary in a version-controlled repository, subject to the source licence.

Author contribution: Nkosana Mkandla conceived the operational problem, developed the ERP capability framework, interpreted the findings and prepared the manuscript.

REFERENCES

[1] Aladwani, A.M. (2001) 'Change management strategies for successful ERP implementation', Business Process Management Journal, 7(3), pp. 266-275. https://doi.org/10.1108/14637150110392764.

[2] Alles, M.G., Kogan, A. and Vasarhelyi, M.A. (2006) 'Continuous monitoring of business process controls: A pilot implementation of a continuous auditing system at Siemens', International Journal of Accounting Information Systems, 7(2), pp. 137-161. https://doi.org/10.1016/j.accinf.2005.10.004.

[3] AlMuhayfith, S. and Shaiti, H. (2020) 'The impact of enterprise resource planning on business performance: With the discussion on its relationship with open innovation', Journal of Open Innovation: Technology, Market, and Complexity, 6(3), 87. https://doi.org/10.3390/joitmc6030087.

[4] Aloini, D., Dulmin, R. and Mininno, V. (2007) 'Risk management in ERP project introduction: Review of the literature', Information & Management, 44(6), pp. 547-567. https://doi.org/10.1016/j.im.2007.05.004.

[5] Amoako-Gyampah, K. and Salam, A.F. (2004) 'An extension of the technology acceptance model in an ERP implementation environment', Information & Management, 41(6), pp. 731-745. https://doi.org/10.1016/j.im.2003.08.010.

[6] Austin, P.C. and Stuart, E.A. (2015) 'Moving towards best practice when using inverse probability of treatment weighting using the propensity score to estimate causal treatment effects in observational studies', Statistics in Medicine, 34(28), pp. 3661-3679. https://doi.org/10.1002/sim.6607.

[7] Bertrand, M., Duflo, E. and Mullainathan, S. (2004) 'How much should we trust differences-in-differences estimates?', Quarterly Journal of Economics, 119(1), pp. 249-275. https://doi.org/10.1162/003355304772839588.

[8] Bharadwaj, A.S. (2000) 'A resource-based perspective on information technology capability and firm performance: An empirical

investigation', MIS Quarterly, 24(1), pp. 169-196. https://doi.org/10.2307/3250983.

[9] Bradford, M. and Florin, J. (2003) 'Examining the role of innovation diffusion factors on the implementation success of enterprise resource planning systems', International Journal of Accounting Information Systems, 4(3), pp. 205-225. https://doi.org/10.1016/S1467-0895(03)00026-5.

[10] Callaway, B. and Sant'Anna, P.H.C. (2021) 'Difference-in-differences with multiple time periods', Journal of Econometrics, 225(2), pp. 200-230. https://doi.org/10.1016/j.jeconom.2020.12.001.

[11] Chingezi, E., Chingezi, L., Ganyani, L., Yelduora, P.G. and Mupa, M.N. (2026a) 'Accounting analytics for inventory integrity and working capital control: Continuous auditing approaches for ERP-enabled supply chains', Iconic Research and Engineering Journals, 10(1), pp. 48-58. https://doi.org/10.64388/IREV10I1-1719387.

[12] Chingezi, E., Chingezi, L., Ganyani, L., Yelduora, P.G. and Mupa, M.N. (2026b) 'Reducing repeat findings and control failures in operationally intensive U.S. enterprises: A data-driven internal audit framework for remediation, loss prevention, and governance improvement', World Journal of Advanced Research and Reviews, 30(3), pp. 2188-2199. https://doi.org/10.30574/wjarr.2026.30.3.1792.

[13] Chou, S.-W. and Chang, Y.-C. (2008) 'The implementation factors that influence the ERP benefits', Decision Support Systems, 46(1), pp. 149-157. https://doi.org/10.1016/j.dss.2008.06.003.

[14] Chukwuemeka, O.D.T. and Mupa, M.N. (2026a) 'Predictive training analytics for competency-based pilot instruction: Early detection of checkride risk, procedural noncompliance and safety-relevant learning gaps', World Journal of Advanced Research and Reviews, 30(3), pp. 1928-1942. https://doi.org/10.30574/wjarr.2026.30.3.1774.

[15] Constante, F., Silva, F. and Pereira, A. (2019) DataCo Smart Supply Chain for Big Data Analysis. Mendeley Data, Version 5. https://doi.org/10.17632/8gx2fvg2k6.5.

[16] Cotteleer, M.J. and Bendoly, E. (2006) 'Order lead-time improvement following enterprise information technology implementation: An empirical study', MIS Quarterly, 30(3), pp. 643-660. https://doi.org/10.2307/25148743.

[17] Davenport, T.H. (1998) 'Putting the enterprise into the enterprise system', Harvard Business Review, 76(4), pp. 121-131.

[18] DeLone, W.H. and McLean, E.R. (2003) 'The DeLone and McLean model of information systems success: A ten-year update', Journal of Management Information Systems, 19(4), pp. 9-30. https://doi.org/10.1080/07421222.2003.11045748.

[19] Eller, R., Alford, P., Kallmunzer, A. and Peters, M. (2020) 'Antecedents, consequences, and challenges of small and medium-sized enterprise digitalization', Journal of Business Research, 112, pp. 119-127. https://doi.org/10.1016/j.jbusres.2020.03.004.

[20] Gattiker, T.F. and Goodhue, D.L. (2005) 'What happens after ERP implementation: Understanding the impact of interdependence and differentiation on plant-level outcomes', MIS Quarterly, 29(3), pp. 559-585. https://doi.org/10.2307/25148695.

[21] Hendricks, K.B., Singhal, V.R. and Stratman, J.K. (2007) 'The impact of enterprise systems on corporate performance: A study of ERP, SCM, and CRM system implementations', Journal of Operations Management, 25(1), pp. 65-82. https://doi.org/10.1016/j.jom.2006.02.002.

[22] Hitt, L.M., Wu, D.J. and Zhou, X. (2002) 'Investment in enterprise resource planning: Business impact and productivity measures', Journal of Management Information Systems, 19(1), pp. 71-98. https://doi.org/10.1080/07421222.2002.11045716.

[23] Kallunki, J.-P., Laitinen, E.K. and Silvola, H. (2011) 'Impact of enterprise resource planning systems on management control systems and firm performance', International Journal of Accounting Information Systems, 12(1), pp. 20-39. https://doi.org/10.1016/j.accinf.2010.02.001.

[24] Kuhn, J.R. and Sutton, S.G. (2010) 'Continuous auditing in ERP system environments: The current state and future directions', Journal of

Information Systems, 24(1), pp. 91-112. https://doi.org/10.2308/jis.2010.24.1.91.

[25] Madapusi, A. and D'Souza, D. (2012) 'The influence of ERP system implementation on the operational performance of an organization', International Journal of Information Management, 32(1), pp. 24-34. https://doi.org/10.1016/j.ijinfomgt.2011.09.001.

[26] McAfee, A. (2002) 'The impact of enterprise information technology adoption on operational performance: An empirical investigation', Production and Operations Management, 11(1), pp. 33-53. https://doi.org/10.1111/j.1937-5956.2002.tb00183.x.

[27] Nah, F.F.-H., Lau, J.L.-S. and Kuang, J. (2001) 'Critical factors for successful implementation of enterprise systems', Business Process Management Journal, 7(3), pp. 285-296. https://doi.org/10.1108/14637150110392782.

[28] Nicolaou, A.I. (2004) 'Firm performance effects in relation to the implementation and use of enterprise resource planning systems', Journal of Information Systems, 18(2), pp. 79-105. https://doi.org/10.2308/jis.2004.18.2.79.

[29] Nicolaou, A.I. and Bhattacharya, S. (2006) 'Organizational performance effects of ERP systems usage: The impact of post-implementation changes', International Journal of Accounting Information Systems, 7(1), pp. 18-35. https://doi.org/10.1016/j.accinf.2005.12.002.

[30] Petter, S., DeLone, W. and McLean, E.R. (2013) 'Information systems success: The quest for the independent variables', Journal of Management Information Systems, 29(4), pp. 7-62. https://doi.org/10.2753/MIS0742-1222290401.

[31] Ram, J., Corkindale, D. and Wu, M.-L. (2013) 'Implementation critical success factors (CSFs) for ERP: Do they contribute to implementation success and post-implementation performance?', International Journal of Production Economics, 144(1), pp. 157-174. https://doi.org/10.1016/j.ijpe.2013.01.032.

[32] Ruivo, P., Oliveira, T. and Neto, M. (2014) 'Examine ERP post-implementation stages of use and value: Empirical evidence from Portuguese SMEs', International Journal of Accounting Information Systems, 15(2), pp. 166-184. https://doi.org/10.1016/j.accinf.2014.01.002.

[33] Seddon, P.B., Calvert, C. and Yang, S. (2010) 'A multi-project model of key factors affecting organizational benefits from enterprise systems', MIS Quarterly, 34(2), pp. 305-328. https://doi.org/10.2307/20721429.

[34] Seethamraju, R. (2013) 'Influence of ERP systems on business process agility', IIMB Management Review, 25(3), pp. 137-149. https://doi.org/10.1016/j.iimb.2013.05.001.

[35] Shang, S. and Seddon, P.B. (2002) 'Assessing and managing the benefits of enterprise systems: The business manager's perspective', Information Systems Journal, 12(4), pp. 271-299. https://doi.org/10.1046/j.1365-2575.2002.00132.x.

[36] Strong, D.M. and Volkoff, O. (2010) 'Understanding organization-enterprise system fit: A path to theorizing the information technology artifact', MIS Quarterly, 34(4), pp. 731-756. https://doi.org/10.2307/25750703.

[37] Sykes, T.A., Venkatesh, V. and Johnson, J.L. (2014) 'Enterprise system implementation and employee job performance: Understanding the role of advice networks', MIS Quarterly, 38(1), pp. 51-72. https://doi.org/10.25300/MISQ/2014/38.1.03.

[38] Trkman, P. (2010) 'The critical success factors of business process management', International Journal of Information Management, 30(2), pp. 125-134. https://doi.org/10.1016/j.ijinfomgt.2009.07.003.

[39] Umble, E.J., Haft, R.R. and Umble, M.M. (2003) 'Enterprise resource planning: Implementation procedures and critical success factors', European Journal of Operational Research, 146(2), pp. 241-257. https://doi.org/10.1016/S0377-2217(02)00547-7.

[40] Uwizeyemungu, S. and Raymond, L. (2012) 'Impact of an ERP system's capabilities upon the realisation of its business value: A resource-based perspective', Information Technology and Management, 13(2), pp. 69-90. https://doi.org/10.1007/s10799-012-0118-9.

[41] Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D. (2003) 'User acceptance of information technology: Toward a unified view', MIS Quarterly, 27(3), pp. 425-478. https://doi.org/10.2307/30036540.

[42] Wang, R.Y. and Strong, D.M. (1996) 'Beyond accuracy: What data quality means to data consumers', Journal of Management

Information Systems, 12(4), pp. 5-33. https://doi.org/10.1080/07421222.1996.11518099.

[43] Xu, H., Nord, J.H., Brown, N. and Nord, G.D. (2002) 'Data quality issues in implementing an ERP', Industrial Management & Data Systems, 102(1), pp. 47-58. https://doi.org/10.1108/02635570210414668.

[44] Zerbino, P., Aloini, D., Dulmin, R. and Mininno, V. (2021) 'Why enterprise resource planning initiatives do succeed in the long run: A case-based causal network', PLOS ONE, 16(12), e0260798. https://doi.org/10.1371/journal.pone.0260798.

Appendix A. Detailed Field KPI Dictionary

Table 19. Publication-ready KPI dictionary for prospective studies

KPI	Definition	Source	Recommended analysis	Control note
Order-to-cash cycle	Calendar days from accepted order to cleared cash	Order, shipment, invoice, receipt timestamps	Median and 90th percentile; separate customer terms	Excludes disputed orders only if reason documented
Inventory accuracy	1 - sum absolute count variance / sum system quantity	Cycle-count and ERP stock records	By site, class and item criticality	Negative or zero denominators require defined handling
Reporting latency	Hours/days from reporting cutoff to approved report availability	Close calendar, workflow audit log	Median by report; 95th percentile	Definition must distinguish draft from approved
Invoice rejection	Rejected or returned invoices / submitted invoices	Billing workflow, customer portal, contract system	By reason, customer/agency and value	Duplicates and administrative returns separated
Delivery reliability	On-time complete deliveries / due deliveries	Promise date, shipment, proof of delivery	By service class, contract and root cause	Promise-date changes logged and governed
User adoption	Active competent users / assigned users	Audit log plus task-based assessment	Role specific; monthly/quarterly	Logins alone are insufficient
Control exception rate	Material exceptions / eligible transactions	Control engine and workflow	By control, cause, owner and recurrence	Control changes versioned
Corrective-action closure	Actions closed with verified evidence / actions due	Issue tracker	On-time and overdue ageing	Reopened actions reported separately
Data-quality defect rate	Validated critical defects / records sampled or processed	Data-quality rules and sampling	Completeness, validity, consistency, uniqueness, timeliness	Automated passes do not replace accuracy sampling

Benefit realisation	Verified cumulative benefit / approved target	Benefit register and finance validation	By benefit owner and intervention component	Avoid double counting and unsupported attribution

Appendix B. ERP Capability Evidence Checklist

- Approved current-state and future-state process maps with end-to-end ownership.
- Requirements traceability linking process, data, control, report and integration needs to testing.
- Master-data standards, ownership, migration reconciliation and unresolved-defect register.
- Role/access matrix, segregation-of-duties analysis and approval-workflow evidence.
- Test coverage for normal, exception, reversal, security, integration and recovery scenarios.
- Role-based training materials, task assessments, attendance and remediation evidence.
- Super-user and support model with severity, response, escalation and closure expectations.
- Baseline KPI pack with signed definitions and historical trends.
- Cutover decision, rollback plan, open-risk acceptance and hypercare plan.
- Post-go-live adoption, data-quality, control, service and benefits dashboard.
- Change-control and release records linking configuration changes to retesting.
- Post-implementation review documenting benefits, failures, lessons and next actions.

Appendix C. Interpretation Guide for the EOCI

Table 20. EOCI interpretation bands

Score	Tier	Recommended response
70-100	Strong	Observed group-level outcomes are comparatively controlled; investigate sustainability, hidden workarounds and cost
60-69.9	Controlled	Generally functioning with targeted weaknesses; prioritise the lowest component and recurrent exceptions
50-59.9	Fragile	Material delivery/control weakness; conduct root-cause review and define corrective programme
Below 50	Critical	Systematic capability mismatch or severe outcome failure; executive ownership and immediate redesign required

Note: Bands are managerial heuristics developed for this paper. They require validation before external benchmarking or high-stakes use.

Appendix D. Reproducibility and Quality-Control Notes

- Use Order Id as the delivery-level aggregation key and verify within-order consistency before aggregation.
- Exclude realised outcomes from predictors when building pre-dispatch risk models.
- Retain a temporally separated or site-separated external test set where available; random holdout is a minimum, not a final validation.
- Report discrimination and calibration together. A high AUC does not guarantee accurate probabilities.
- Version all metric definitions, data transformations, exclusion rules, model code and dashboard logic.
- Preserve a cohort-flow table showing source rows, exclusions, duplicates, aggregation and final samples.
- Assess missingness and data-quality mechanisms; do not interpret complete fields as necessarily accurate.
- Perform subgroup and site-level analyses before operational deployment, and document where estimates are unstable.
- Use human review for exceptions and record overrides, rationales and outcomes so that the workflow itself can be evaluated.
- Reproduce the present analysis on the full source dataset before journal submission and archive the exact extract checksum.