© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

A Practical Digital-Transformation Maturity and Delivery-Readiness Framework for U.S. SMEs and Small Government Vendors

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* - *Digital transformation in small and medium-sized enterprises (SMEs) is frequently measured by technology adoption rather than by the operating capability created after adoption. This study develops a practical Digital-Transformation Maturity and Delivery-Readiness Framework for U.S. SMEs and small government vendors. The framework integrates seven domains: process integration, ERP readiness, data quality, internal controls, cyber hygiene, user adoption and delivery capability. A 28-item instrument was constructed through design-science synthesis of digital transformation, information-systems success, ERP, data-governance, control, cybersecurity and maturity-model research. Pre-pilot validation used a fully disclosed synthetic panel of 720 hypothetical firms and an external operational benchmark derived from 70,000 DataCo supply-chain records covering 44,026 unique orders and 229 market-department-segment-period units. The synthetic panel was designed to test factor recovery, reliability, discriminant validity, scoring sensitivity and predictive behavior before human-subject deployment. Kaiser-Meyer-Olkin adequacy was 0.936; Bartlett's test was significant (chi-square=9398.6, df=378, p<0.001); and the intended seven-factor structure explained 67.4% of item variance. Construct alpha coefficients ranged from 0.817 to 0.852, composite reliability from 0.880 to 0.901, and average variance extracted from 0.647 to 0.694. The maximum heterotrait-monotrait ratio was 0.639. Five-fold cross-validated logistic regression predicted implementation readiness with AUC=0.799, while linear regression predicted delivery performance with R-squared=0.784. Index rankings remained highly stable across equal, outcome-derived and risk-balanced weights (Spearman rho at least 0.995). Maturity stages showed monotonic increases in readiness and reliable delivery. The contribution is an instrument, scoring architecture, risk heat-map method, benchmark dashboard and staged roadmap that can be tested prospectively. The paper explicitly limits inference: the findings establish pre-pilot measurement plausibility, not population prevalence or causal impact. A multi-site U.S. validation protocol is specified for the next phase.*

Keywords: digital transformation; maturity model; ERP readiness; SME; government vendor; data quality; internal controls; cyber hygiene; user adoption; delivery readiness; psychometric validation

I. INTRODUCTION

Digital transformation is not equivalent to digitisation, software acquisition or isolated automation. It is an organisational change process in which digital technologies reshape value creation, decision rights, operating routines and relationships with customers, suppliers and public institutions (Vial, 2019; Verhoef et al., 2021). This distinction is especially important for SMEs. Smaller firms can acquire cloud applications quickly, but they often lack the managerial slack, specialist staff, formal governance and process discipline required to translate adoption into repeatable performance (Eller et al., 2020; Garzoni et al., 2020).

The practical failure mode is familiar: an organisation purchases an ERP or cloud platform, migrates partial data, trains users on screens rather than roles, preserves manual workarounds and declares implementation complete at go-live. The system exists, yet management still cannot trust master data, trace approvals, reconcile exceptions, predict delivery failure or demonstrate contract performance. Enterprise-system research has long shown that technical implementation is only one part of performance realisation; process fit, data integrity, organisational learning and post-implementation use are decisive (Davenport, 1998; Gattiker and Goodhue, 2005; Ruivo, Oliveira and Neto, 2014).

The U.S. setting makes the measurement problem consequential. Small businesses constitute the overwhelming majority of U.S. firms, and many operate with limited internal technology capacity. NIST reported 34.8 million small businesses and noted that more than four-fifths have no paid employees beyond their owners, illustrating the heterogeneity and resource constraints within the sector (Eliot, Marron and Thorn, 2026). International survey evidence corroborates this picture: the OECD's 2025 D4SME survey finds that SME digital adoption is highly uneven across core business functions and that capability gaps, not access to technology, are the binding constraint for most firms (Bianchini and Lasheras Sancho, 2025). Small government vendors face an additional burden: they must combine commercial agility with disciplined scheduling, documentation, invoice support, access control, subcontractor oversight and corrective-action closure. A maturity framework intended for this population must therefore connect technology with delivery assurance rather than score technology possession alone.

Existing digital-maturity models provide valuable taxonomies, but many are enterprise-centric, consultancy-driven, weakly validated or overly linear, echoing longstanding concerns in the broader technology-adoption review literature that firm-level adoption models are frequently applied without adequate attention to the SME context in which they are deployed (Oliveira and Martins, 2011; Ifinedo, 2011). Maturity scores may be attractive as dashboards while concealing three methodological weaknesses: unclear construct definitions, arbitrary weights and limited evidence that higher scores predict implementation or delivery outcomes (Becker, Knackstedt and Poppelbuss, 2009; Mettler, 2011). A useful SME instrument must be short enough to administer, explicit enough to audit, multi-domain enough to detect bottlenecks and statistically testable.

This study addresses that gap by developing a seven-domain Digital-Transformation Maturity and Delivery-Readiness Framework (DTM-DRF). The framework treats maturity as a configuration of complementary capabilities rather than a single technology ladder. It links process integration, ERP readiness, data quality, internal controls, cyber hygiene, user adoption and delivery capability to two outcomes: implementation readiness and operational delivery performance. It also produces operational outputs required by practitioners: an assessment instrument, maturity stages, risk heat maps, a benchmark dashboard and an implementation roadmap.

The empirical strategy is intentionally staged. A transparent synthetic panel is used to stress-test psychometric properties and scoring behavior before collecting human-subject data. A public DataCo supply-chain extract is used to calibrate the external distribution of operational delivery performance and to demonstrate cross-segment heat maps. This combination supports measurement engineering while avoiding the false claim that simulated records are observed firms. Full validation is reserved for the consenting U.S. multi-site protocol specified later in the paper.

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, "From Software Acquisition to Operational Capability: Measuring ERP-Enabled Performance," analyses the same extract at transaction level to benchmark delivery, margin and workflow outcomes and to develop a leakage-controlled readiness model and an ERP Operational Capability Index. The present paper instead uses the extract only to calibrate the external distribution of a delivery-performance score for a distinct seven-domain organisational maturity instrument tested on a separate synthetic panel. The two studies report order-level statistics from the same public source (for example, the late-delivery rate) and are reconciled to the same order-level computation to avoid presenting inconsistent descriptions of shared source data; neither paper's synthetic, predictive or index results depend on the other.

1.1 Research question and objectives

Primary research question
Which maturity dimensions predict implementation readiness and delivery performance among U.S. SMEs and small government vendors?

1. Develop conceptually distinct, auditable dimensions of digital-transformation maturity

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

relevant to resource-constrained firms and small public-sector vendors.
2. Construct a concise 28-item instrument and scoring method that produces dimension scores, a composite index, maturity stages and risk flags.
3. Pre-test reliability, factor recovery, convergent validity, discriminant validity and predictive behavior in a transparent synthetic validation panel.
4. Calibrate delivery-performance scoring to a public operational dataset and demonstrate benchmark and risk-heat-map outputs.
5. Specify expert-review, consenting pilot-survey, case-validation and longitudinal outcome protocols for field deployment.

1.2 Hypotheses

H1: Process integration is positively associated with implementation readiness and delivery performance.
H2: ERP readiness is positively associated with implementation readiness and delivery performance.
H3: Data quality is positively associated with implementation readiness and delivery performance.
H4: Internal controls are positively associated with implementation readiness and delivery performance.
H5: Cyber hygiene is positively associated with implementation readiness and delivery performance.
H6: User adoption is positively associated with implementation readiness and delivery performance.
H7: Delivery capability is positively associated with delivery performance and contributes to implementation-readiness classification.
H8: The composite maturity index is robust to reasonable alternative weighting schemes.
H9: Higher maturity stages exhibit monotonic increases in implementation readiness and reliable delivery.

II. LITERATURE REVIEW AND THEORETICAL FOUNDATION

2.1 Digital transformation as organisational capability

Digital transformation research has moved beyond the idea that performance follows automatically from technology investment. Vial (2019) defines digital transformation through disruptions, strategic responses, structural changes and value-creation paths. Verhoef et al. (2021) similarly distinguish digitisation, digitalisation and transformation, while Warner and Wager (2019) emphasise dynamic capabilities for sensing, seizing and transforming. For SMEs, digital change is shaped by owner-manager commitment, resource scarcity, ecosystem dependence and the ability to reconfigure routines (Li et al., 2018; Eller et al., 2020).

The resource-based view and dynamic-capabilities perspective explain why identical software can produce different outcomes. Technology becomes strategically valuable when combined with complementary organisational resources that are difficult to imitate, including process knowledge, governance routines and learning capacity (Barney, 1991; Teece, 2007). Sambamurthy, Bharadwaj and Grover (2003) connect digital options with agility, while Matarazzo et al. (2021) show how dynamic capabilities mediate digital transformation and customer value creation in SMEs. A maturity instrument should therefore assess capability combinations, not application counts.

Business-model research also indicates that digital investments create value when firms adapt how they organise and transact. Bouwman, Nikou and de Reuver (2019) link business-model experimentation and digital strategy to SME performance. Priyono, Moin and Putri (2020) identify multiple transformation pathways rather than one universal sequence. These findings support a staged but non-deterministic framework: firms can have different profiles at the same overall score, and the lowest capability may constrain the whole system.

2.2 ERP readiness, process integration and information-systems success

ERP readiness is more than vendor selection. It includes a defined business case, process ownership, fit-gap analysis, implementation capacity, migration planning, testing, cutover discipline and benefits governance. ERP risk research identifies inadequate change management, poor requirements, data problems, excessive customisation and weak project governance as recurring failure sources (Aloini, Dulmin and Mininno, 2007; Umble, Haft and Umble, 2003). Post-implementation research shows that continued use, integration and organisational learning determine whether firms realise value after go-live (Gattiker and Goodhue, 2005; Ruivo, Oliveira and Neto, 2014).

Process integration represents the end-to-end coherence of workflows and handoffs. Business-process management research emphasises strategic alignment, governance, methods, information technology, people and culture (Trkman, 2010). ERP can standardise transactions, but it can also encode fragmentation if the future-state process is not designed first. DTM-DRF therefore separates ERP readiness from process integration: a firm may be project-ready without having integrated processes, or process-disciplined while using modest technology.

The DeLone and McLean (2003) information-systems success model distinguishes system quality, information quality, service quality, use, user satisfaction and net benefits. This structure informs the present framework but is extended in two ways. First, readiness and delivery are treated as outcomes requiring governance and controls. Second, the model is adapted to the operational realities of smaller firms and vendors whose performance is judged through schedule, documentation, invoice and corrective-action evidence.

2.3 Data quality, internal controls and cyber hygiene

Data quality is a multidimensional business property rather than a technical cleaning task. Wang and Strong (1996) identify intrinsic, contextual, representational and accessibility dimensions; Xu et al. (2002) demonstrate the importance of data-quality issues in ERP implementation. Master-data ownership, validation rules, reconciliation, lineage and timeliness affect planning, invoicing, inventory and management reporting. Data quality is therefore measured independently rather than absorbed into system quality.

Internal controls convert digital workflows into reliable evidence. The COSO framework organises control environment, risk assessment, control activities, information and communication, and monitoring (COSO, 2013). Recent applied SME research has highlighted how weak policy formalisation, manual accounting and fragmented data create governance debt, and how analytics can support audit readiness and remediation (Yelduora et al., 2026). Continuous ERP analytics can identify inventory-integrity, margin and working-capital exceptions, but analytics only becomes governance when ownership, escalation and closure are explicit (Chingezi et al., 2026b).

Cyber hygiene is included because integrated operations expand dependence on identities, endpoints, vendors, backups and data exchange. NIST CSF 2.0 is organised around Govern, Identify, Protect, Detect, Respond and Recover and is designed for organisations of different sizes and maturity (Pascoe, Quinn and Scarfone, 2024). NIST SP 1300 adapts those principles for resource-constrained small businesses (Eliot, 2024). Cyber maturity is not treated as a substitute for specialised assessment; the instrument captures minimum governance and operational hygiene relevant to digital readiness.

2.4 User adoption, training and delivery capability

User adoption is a behavioural and organisational capability. UTAUT identifies performance expectancy, effort expectancy, social influence and facilitating conditions as determinants of technology use (Venkatesh et al., 2003). Technology readiness adds individual propensities toward new technologies (Parasuraman, 2000). For operational transformation, however, attendance at training is not enough. Firms need role-based proficiency, super-user support, usage monitoring, issue closure and a feedback loop that improves both process and system.

A useful methodological lesson from competency-based analytics generally is that training data should detect procedural noncompliance and learning gaps early enough for intervention, rather than merely document completion. Applied SME audit research makes a parallel move from retrospective reporting to decision intelligence and continuous anomaly detection (Taanisa et al., 2026a). DTM-DRF therefore scores adoption through evidenced use and learning, not course completion alone.

Delivery capability is the bridge between internal maturity and external reliability. For SMEs and small government vendors, it includes schedule and milestone control, order-to-cash and invoice documentation, quality and delivery evidence, subcontractor coordination and corrective-action closure. Delivery capability is modelled both as a maturity dimension and a direct predictor of operational performance. This dual role reflects its position as an organisational capability that is also visible in outcomes.

2.5 Maturity-model design and validation

Maturity models translate complex capability development into diagnostic stages. Their usefulness depends on a defined purpose, explicit constructs, assessment procedures, improvement logic and evidence of validity (Becker, Knackstedt and Poppelbuss, 2009; Poppelbuss and Roglinger, 2011). Mettler (2011) argues that maturity assessment

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

should be treated as design-science research rather than as an untested checklist.

Scale-development research requires a domain definition, item generation, content review, purification and validation (Churchill, 1979; Hinkin, 1998; MacKenzie, Podsakoff and Podsakoff, 2011). Internal consistency alone is insufficient. Convergent validity, discriminant validity, factor structure and predictive validity must also be examined. Cronbach alpha, composite reliability, average variance extracted and HTMT are used here as complementary diagnostics (Cronbach, 1951; Fornell and Larcker, 1981; Henseler, Ringle and Sarstedt, 2015).

Table 1. DTM-DRF domains, definitions and auditable evidence

Domain	Operational definition	Illustrative evidence
Process integration (PI)	End-to-end standardisation, cross-functional handoffs, system integration and exception learning	Workflow maps; handoff controls; integration logs; rework metrics
ERP readiness (ER)	Business case, fit-gap requirements, implementation capacity, migration/testing and cutover readiness	Approved charter; process ownership; test evidence; cutover plan
Data quality (DQ)	Ownership, validation, completeness, accuracy, timeliness, lineage and reconciliation	Data dictionary; quality rules; reconciliations; issue ageing
Internal controls (IC)	Access, segregation, approvals, evidence, monitoring and remediation closure	Role matrix; approval logs; reconciliations; action tracker
Cyber hygiene (CH)	Governance, identity protection, backup/recovery, vulnerability and incident preparedness	Asset/account inventory; MFA; recovery test; incident procedure
User adoption (UA)	Role proficiency, super-user support, observed use, feedback and continuous learning	Competency tests; usage logs; help-desk trends; adoption reviews
Delivery capability (DC)	Schedule, cost, invoice, quality, subcontractor, documentation and corrective-action performance	Milestone plan; invoice pack; delivery evidence; corrective-action closure

Note: Each domain contains four core items scored from 1 (absent/ad hoc) to 5 (embedded, measured and improved).

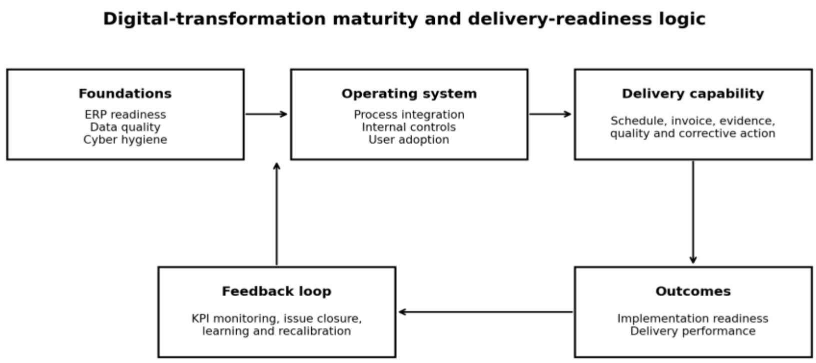

Figure 1. Conceptual logic of digital-transformation maturity and delivery readiness

The feedback loop makes maturity dynamic: operational outcomes inform issue closure, learning and recalibration.

III. FRAMEWORK AND INSTRUMENT DEVELOPMENT

3.1 Design principles

6. Capability rather than possession: the instrument asks whether practices are evidenced, repeatable and measured, not whether software has been purchased.
7. Multi-domain complementarity: process, technology, data, controls, cyber, people and delivery are assessed separately to expose bottlenecks.
8. Auditable scoring: every rating should be supported by observable evidence, an accountable owner and an assessment date.
9. Proportionality: the core instrument is usable by resource-constrained firms; advanced controls can be added without changing the basic scale.
10. Outcome linkage: scores are evaluated against implementation readiness and delivery performance rather than treated as self-validating.
11. Uncertainty and transparency: synthetic and public-data results are labelled accurately, alternative weights are tested and causal claims are reserved for longitudinal field data.

3.2 Item construction and response scale

Four items were developed for each of seven domains, producing 28 core items. Items were written as organisational practices that can be rated by a knowledgeable respondent and corroborated through evidence. The proposed response anchors are: 1=absent or purely ad hoc; 2=initial and person-dependent; 3=defined but inconsistently applied; 4=controlled and routinely evidenced; and 5=embedded, measured and continuously improved. Domain scores are transformed to a 0-100 scale using (mean item score - 1)/4 x 100.

Table 2. Core 28-item assessment instrument

Item	Assessment statement	Scale
PI1	End-to-end workflows are documented and owned across functions.	1-5
PI2	Handoffs and exception routes are standardised.	1-5
PI3	Core systems exchange data without repeated manual re-entry.	1-5
PI4	Rework, queue time and process exceptions are measured and reviewed.	1-5
ER1	A quantified business case, scope and process ownership are approved.	1-5
ER2	Fit-gap requirements and vendor-neutral acceptance criteria are documented.	1-5
ER3	Resources, change impacts, risks and decision rights are established.	1-5
ER4	Data migration, testing, cutover and stabilisation plans are ready and rehearsed.	1-5
DQ1	Master-data domains have named owners and definitions.	1-5

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

Item	Assessment statement	Scale
DQ2	Validation rules address duplicates, invalid values and missing critical fields.	1-5
DQ3	Completeness, accuracy, consistency and timeliness are monitored.	1-5
DQ4	Source-to-report lineage and reconciliations are documented.	1-5
IC1	Role-based access and segregation-of-duties risks are reviewed.	1-5
IC2	Approvals, thresholds and supporting evidence are enforced.	1-5
IC3	Reconciliations and exception monitoring occur at defined frequencies.	1-5
IC4	Control failures have owners, due dates, root causes and verified closure.	1-5
CH1	Critical assets, accounts and data are inventoried; strong authentication is used.	1-5
CH2	Backups are protected and recovery is tested.	1-5
CH3	Patching, vulnerability handling and incident escalation are defined.	1-5
CH4	Third-party risk and workforce security awareness are reviewed.	1-5
UA1	Training is role-based and verifies practical proficiency.	1-5
UA2	Super users and support routes are available after go-live.	1-5
UA3	Actual system use, workarounds and process compliance are measured.	1-5
UA4	User feedback, support trends and lessons learned drive improvement.	1-5
DC1	Schedule, cost, milestones and delivery commitments are actively controlled.	1-5
DC2	Order-to-cash and invoice documentation are complete and timely.	1-5

Item	Assessment statement	Scale
DC3	Quality, delivery and corrective-action performance are monitored.	1-5
DC4	Subcontractor, contract and evidence obligations are traceable and current.	1-5

Note: Appendix A provides scoring guidance and evidence examples. The instrument is intended for multi-respondent administration where feasible.

3.3 Composite score, stages and bottleneck logic

The base index is a weighted average of the seven 0-100 domain scores. The pre-pilot outcome-derived weights are PI 16.6%, ER 16.4%, DQ 15.7%, IC 16.2%, CH 6.7%, UA 13.7% and DC 14.7%. These weights are provisional because they were estimated in synthetic data. Equal and risk-balanced alternatives are reported in sensitivity analysis. For field use before external validation, the risk-balanced scheme is recommended because it avoids overfitting and preserves cyber hygiene as a minimum capability.

Table 3. Maturity stages and primary management response

Stage	Index range	Interpretation	Priority action
1 Fragmented	0-39.9	Person-dependent, reactive and weakly evidenced	Stabilise critical workflows, identities, backups and ownership.
2 Emerging	40-54.9	Defined practices exist but are inconsistent or siloed	Standardise processes, data rules, controls and role training.
3 Integrated	55-69.9	Cross-functional practices are repeatable and increasingly measured	Close integration gaps, formalise benefits tracking and delivery dashboards.
4 Governed	70-84.9	Practices are controlled, evidenced and management-reviewed	Strengthen predictive monitoring, third-party assurance and continuous improvement.
5 Adaptive	85-100	Capabilities are measured, learning-driven and resilient	Benchmark externally, automate control monitoring and diffuse proven practices.

Note: A high composite score does not override a critical-domain floor. Any domain below 40 triggers a red bottleneck flag; any cyber or internal-control score below 40 blocks a fully governed rating.

IV. DATA AND METHODS

4.1 Research design

The study follows a design-science and measurement-development sequence: problem definition; construct synthesis; item design; transparent synthetic pre-pilot testing; public operational benchmark calibration; scoring sensitivity; representative case analysis; and specification of expert and field validation. The purpose of synthetic testing is computational quality assurance. It can reveal whether a proposed structure is recoverable and whether a scoring algorithm behaves coherently, but it cannot substitute for observed respondents.

4.2 Public operational benchmark

The public benchmark is a 70,000-record analytical extract of the DataCo SMART Supply Chain dataset, a structured supply-chain dataset created by Constante, Silva and Pereira (2019) and mirrored on Kaggle. The extract covers 44,026 unique orders and includes actual and scheduled shipping days, delivery status, late-delivery risk, order status, profit, customer segment, market, department and shipping mode. Records were grouped into market x department x customer-segment x year-quarter units, retaining units with at least 60 transactions; 229 units remained. The source dataset is not a survey of U.S. SMEs or government vendors. Its role is limited to operational calibration and demonstration of delivery-risk heat maps.

Table 4. Operational delivery benchmark score

Component	Weight (%)	Operationalisation
On-time delivery	32	1 - late-delivery rate
Avoidance of severe delay	18	1 - rate of actual shipping at least two days beyond schedule
Workflow integrity	16	1 - suspected fraud, cancellation, pending-payment or on-hold rate
Margin integrity	14	1 - negative-profit rate
Data completeness	10	Proportion of critical operational fields populated
Schedule conformance	10	Penalty for positive actual-versus-scheduled shipping gap

Note: The 0-100 operational score is a transparent calibration device, not a validated universal performance index.

The component weights in Table 4 were set judgmentally rather than estimated from an external criterion, prioritising outcomes with the most direct customer, contract and cash-flow consequence: on-time delivery (32%) and avoidance of severe delay (18%) receive the largest shares because missed or badly slipped deliveries carry the most immediate commercial and contractual risk; workflow integrity (16%) and margin integrity (14%) follow because they affect solvency and control exposure; data completeness and schedule conformance (10% each) are weighted lowest because they are treated as minimum conditions rather than differentiators. This judgmental basis is a limitation, not a finding, and it is the reason Section 5.8 subjects the maturity-index weights, but not yet this benchmark weighting, to sensitivity testing; a future revision should apply the same rank-correlation sensitivity check used in Table 13 to Table 4's weights, and should treat the DataCo extract's near-even late-delivery split as a property of this particular international e-commerce source rather than a general external prior for U.S. SME or government-vendor delivery performance.

4.3 Synthetic pre-pilot panel

A synthetic panel of 720 hypothetical firms was generated from seven correlated latent capabilities. Four ordinal items per construct were produced on the 1-5 scale with target loadings between approximately 0.73 and 0.84. Context variables represented micro, small, medium and upper-medium firms; six industries; four U.S. regions; years operating; and a government-vendor indicator. Limited contextual shifts were introduced to reflect plausible differences in control and cyber capacity by size and in documentation burden among vendors. No synthetic record corresponds to an identifiable business.

Implementation readiness was generated probabilistically from ERP readiness, process integration, data quality, internal controls, adoption and contextual factors. Delivery performance was generated from delivery capability, process integration, adoption, data quality, controls, cyber hygiene and a weakest-link penalty, then rank-mapped to the empirical DataCo operational-score distribution. This calibration preserves an externally observed performance scale while retaining full disclosure that the relationship between maturity and outcomes is simulated.

Table 5. Synthetic pre-pilot panel characteristics

Variable	Category	n	Percent
Firm size	Medium (50-249)	231	32.1%
Firm size	Small (20-49)	225	31.2%
Firm size	Micro (5-19)	204	28.3%
Firm size	Upper-medium (250-499)	60	8.3%
Industry	Retail/wholesale	155	21.5%
Industry	Professional services	153	21.2%
Industry	Manufacturing	127	17.6%
Industry	Technology	106	14.7%
Industry	Construction	99	13.8%
Industry	Health/social services	80	11.1%
Region	South	260	36.1%
Region	West	169	23.5%
Region	Midwest	167	23.2%
Region	Northeast	124	17.2%
Government vendor	0	467	64.9%
Government vendor	1	253	35.1%

Note: These distributions are design choices for stress testing and must not be interpreted as estimates of the U.S. business population.

4.4 Psychometric analysis

Internal consistency was examined with Cronbach alpha. Composite reliability and average variance extracted (AVE) were estimated from one-factor loadings within each domain. Sampling adequacy was assessed with the Kaiser-Meyer-Olkin statistic, and Bartlett's test evaluated whether the item correlation matrix differed from an identity matrix. Seven-factor exploratory recovery used principal components with varimax rotation, with factor columns aligned to the intended item blocks. Discriminant validity was assessed using the heterotrait-monotrait (HTMT) ratio. Conventional interpretive thresholds were used as guides rather than automatic pass/fail rules: alpha and composite reliability near or above 0.70, AVE above 0.50 and HTMT below 0.85 (Fornell and Larcker, 1981; Henseler, Ringle and Sarstedt, 2015; Hair et al., 2019).

4.5 Predictive and sensitivity analysis

Implementation readiness was modelled through robust logistic regression using standardised domain scores. Delivery performance was modelled through OLS with HC3 robust standard errors. Five-fold cross-validation compared logistic regression with random forest classification using AUC, Brier score, accuracy, F1, calibration intercept and calibration slope. Linear and random-forest regression were compared using RMSE, MAE and R-squared. Random-forest feature importance provided a nonlinear stability check. Three index weights were compared: equal, outcome-derived and risk-balanced. Sensitivity metrics included Spearman rank correlation, mean absolute score shift, readiness AUC and correlation with delivery performance.

4.6 Reproducibility, ethics and assurance

All results are reproducible from the stated random seed, the public DataCo extract and the documented transformation rules. The study uses no personal or confidential human-subject data. Synthetic records

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

are clearly marked and excluded from population inference. Before field deployment, the instrument should undergo independent expert review, cognitive interviewing, ethics review where required, informed consent, multi-respondent reliability testing and preregistered analysis, with the analysis plan registered on a public registry such as OSF prior to data collection. Because the author does not currently hold a university appointment, ethics review is anticipated to be secured either through affiliation with a PhD-granting institution at enrolment or through a commercial/independent IRB; this dependency should be resolved before pilot recruitment begins. This assurance logic follows the broader principle that model discrimination, calibration, workflow validity, subgroup performance and governance should be assessed separately rather than collapsed into a single claim.

V. RESULTS

5.1 Public operational benchmark

The DataCo extract contained 70,000 transaction lines and 44,026 unique orders. The overall late-delivery rate was 54.7% (order-level; see companion note below on cross-study reconciliation). After aggregation and minimum-volume filtering, the mean operational delivery score across 229 units was 69.15 (SD=2.40). The relatively narrow unit-level distribution is useful for calibration but also cautions against treating small score differences as substantive without uncertainty intervals and repeated measurement. This figure is computed at order level and is reported here to be directly comparable with the order-level rate reported in the companion benchmarking study (see Section 1 note on related work); the extract's own late-delivery indicator is close to an even split, which is itself a property of the source dataset that a reader should weigh before treating the 32%-weighted on-time-delivery component in Table 4 as a general external prior.

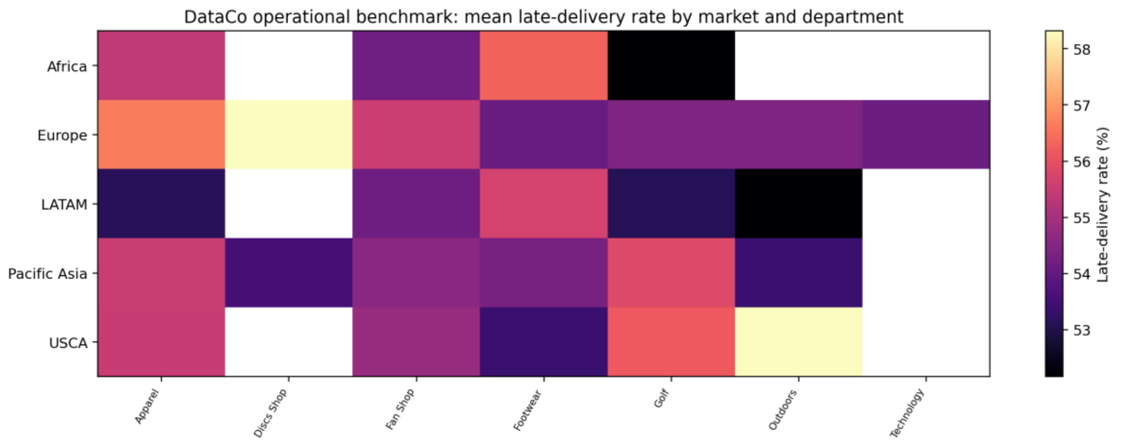

Figure 2. Late-delivery risk by market and department in the public DataCo benchmark

Cells show mean late-delivery percentages for eligible operational units. Blank cells indicate insufficient or unavailable observations. The heat map demonstrates benchmarking architecture, not U.S. vendor prevalence.

The heat map demonstrates why a maturity assessment should not stop at a composite score. Similar enterprise-wide averages can conceal concentrated delivery risk in particular market-department combinations. In field deployment, the same logic should be applied to customer segment, contract, project, process owner, shipping configuration and root-cause category, subject to minimum cell counts and privacy controls.

5.2 Construct reliability and convergent validity

Table 6. Pre-pilot reliability and convergent-validity results

Construct	Alpha	CR	AVE	Loading range	Mean /100
Process integration	0.828	0.886	0.661	0.782-0.841	51.7

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

Construct	Alpha	CR	AVE	Loading range	Mean /100
ERP readiness	0.836	0.891	0.672	0.803-0.836	49.8
Data quality	0.817	0.880	0.647	0.762-0.847	51.1
Internal controls	0.847	0.898	0.687	0.797-0.861	51.9
Cyber hygiene	0.838	0.892	0.675	0.785-0.847	51.0
User adoption	0.824	0.884	0.655	0.783-0.837	50.1
Delivery capability	0.852	0.901	0.694	0.813-0.847	50.3

Note: Results derive from transparent synthetic responses and show designed measurement behavior, not field estimates.

All seven constructs exceeded conventional pre-pilot reliability and convergent-validity guidelines. Alpha ranged from 0.817 to 0.852; composite reliability from 0.880 to 0.901; and AVE from 0.647 to 0.694. These results indicate that the four items within each designed domain cohere without being identical. Because the data-generating process was structured to contain the seven domains, the results are a test of implementation integrity rather than independent empirical confirmation.

5.3 Factor recovery and discriminant validity

The KMO statistic was 0.936, and Bartlett's test was significant (chi-square=9398.6, df=378, p<0.001). The seven rotated factors explained 67.4% of item variance. The loading heat map recovered the intended block structure, with high loadings on each item's target domain and substantially lower cross-loadings.

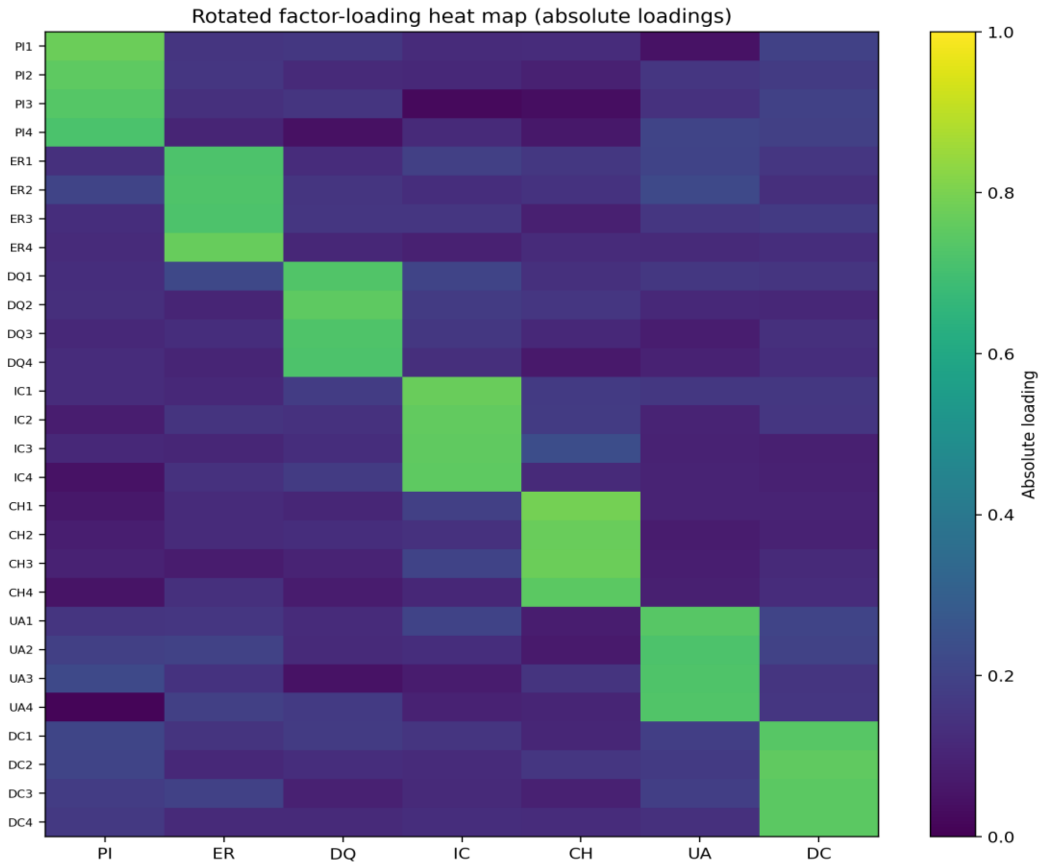

Figure 3. Rotated factor-loading heat map for the 28-item instrument

Absolute loadings are shown. The clear block structure is expected in a designed pre-pilot panel and should be retested through exploratory and confirmatory factor analysis in independent field samples.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

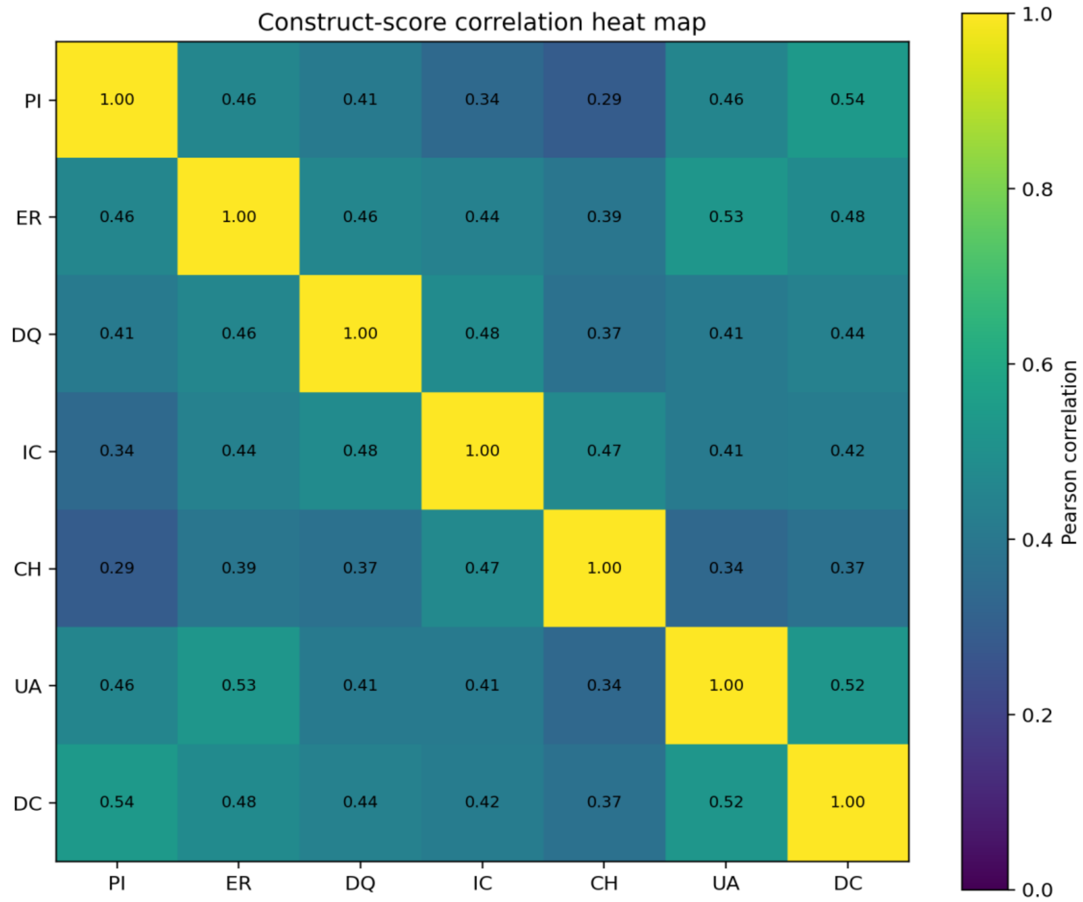

Figure 4. Construct-score correlation heat map

Correlations reflect complementary capabilities while retaining discriminant separation. Maximum HTMT=0.639.

The maximum HTMT ratio was 0.639, below the conservative 0.85 guideline. This supports pre-pilot separation among the seven domains. In a field sample, discriminant validity will be more difficult because common-method variance, respondent halo effects and genuine organisational complementarity can raise inter-construct correlations. Multi-respondent and evidence-based scoring are therefore integral to the validation protocol.

5.4 Predictors of implementation readiness

Table 7. Robust logistic regression for implementation readiness

Domain	Odds ratio per 1 SD	95% CI	p
PI	1.34	1.06-1.68	0.013
ER	1.76	1.39-2.23	<0.001
DQ	1.45	1.16-1.82	<0.001
IC	1.61	1.29-2.00	<0.001
CH	1.17	0.94-1.44	0.153
UA	1.19	0.93-1.51	0.161
DC	1.02	0.82-1.28	0.838

Note: All domain scores were entered jointly. Odds ratios therefore represent conditional associations within the designed synthetic panel.

ERP readiness was the strongest conditional readiness predictor (OR=1.76 per standard deviation), followed by internal controls (OR=1.61), data quality (OR=1.45) and process integration (OR=1.34). Cyber hygiene and user adoption were positive but not statistically significant after adjustment, while delivery capability contributed little to the readiness equation. This pattern is coherent with the outcome design: readiness is primarily an implementation-governance construct, whereas delivery capability becomes more influential in operational performance.

5.5 Predictors of delivery performance

Table 8. Robust OLS model for delivery performance

Domain	Score change per 1 SD	95% CI	p
PI	0.536	0.431-0.641	<0.001
ER	0.149	0.049-0.249	0.004
DQ	0.368	0.270-0.466	<0.001
IC	0.260	0.162-0.359	<0.001
CH	0.169	0.078-0.260	<0.001
UA	0.535	0.429-0.641	<0.001
DC	0.794	0.685-0.904	<0.001

Note: The dependent variable is the 0-100 delivery score calibrated to the public operational distribution.

Delivery capability had the largest adjusted coefficient (0.794 points per standard deviation), followed by process integration (0.536), user adoption (0.535), data quality (0.368), internal controls (0.260), cyber hygiene (0.169) and ERP readiness (0.149). The result reinforces the framework's central proposition: implementation readiness and delivery performance are related but not interchangeable. A firm can be ready to implement and still fail to realise delivery benefits if adoption and delivery controls remain weak.

Table 9. Random-forest feature-importance stability check

Domain	Readiness importance	Delivery importance
DC	0.077	0.508
UA	0.104	0.170
PI	0.149	0.146
DQ	0.164	0.058
ER	0.237	0.055
IC	0.153	0.038
CH	0.117	0.026

Note: Impurity-based importance is descriptive and may favour variables with stronger nonlinear signal. It is used only as a stability check alongside regression coefficients.

5.6 Cross-validated predictive performance and calibration

Table 10. Five-fold cross-validated readiness classification

Model	AUC	Brier	Accuracy	F1	Cal. int.	Cal. slope
Logistic regression	0.799	0.183	0.724	0.736	0.012	0.880

Model	AUC	Brier	Accuracy	F1	Cal. int.	Cal. slope
Random forest	0.796	0.185	0.722	0.734	0.067	0.972

Note: Ideal calibration has intercept 0 and slope 1. Performance is evaluated out of fold.

Table 11. Five-fold cross-validated delivery prediction

Model	RMSE	MAE	R-squared
Linear regression	1.092	0.851	0.784
Random forest	1.192	0.922	0.743

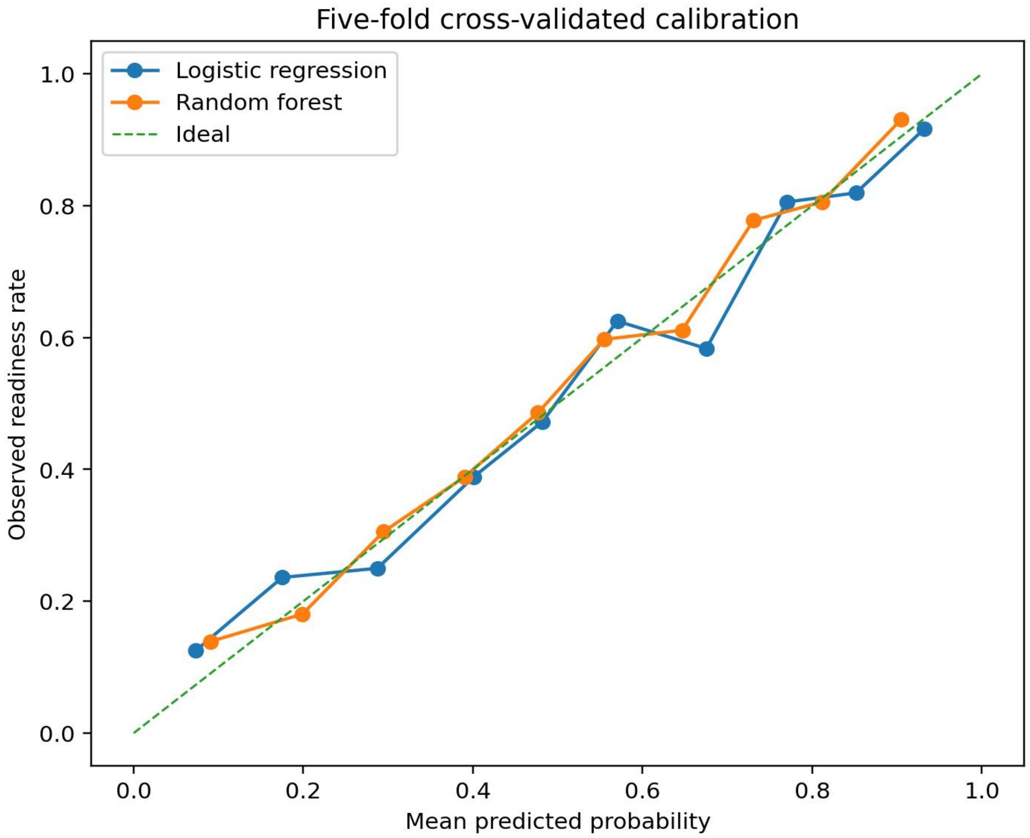

Figure 5. Calibration of readiness predictions

Observed readiness is compared with predicted probability across bins. The figure is a pre-pilot diagnostic, not a field-deployment performance claim.

Logistic regression achieved AUC=0.799 and the random forest AUC=0.796. Their Brier scores were similar. For delivery prediction, linear regression achieved R-squared=0.784, exceeding the random forest's 0.743. The absence of a material nonlinear advantage supports use of a transparent weighted index at the pre-pilot stage. Field data may reveal interactions, thresholds and sector effects that justify more flexible models.

5.7 Maturity stages and outcome gradients

Table 12. Maturity-stage outcome profile

Stage	n	Mean index	Ready	Delivery score	Reliable delivery
1 Fragmented	143	32.6	16.1%	66.31	4.2%

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

Stage	n	Mean index	Ready	Delivery score	Reliable delivery
2 Emerging	306	47.9	44.1%	68.67	43.5%
3 Integrated	225	61.3	77.8%	70.78	93.8%
4 Governed	43	74.8	93.0%	73.07	100.0%
5 Adaptive	3	88.7	100.0%	75.51	100.0%

Note: Adaptive-stage results contain only three synthetic cases and should not be interpreted precisely.

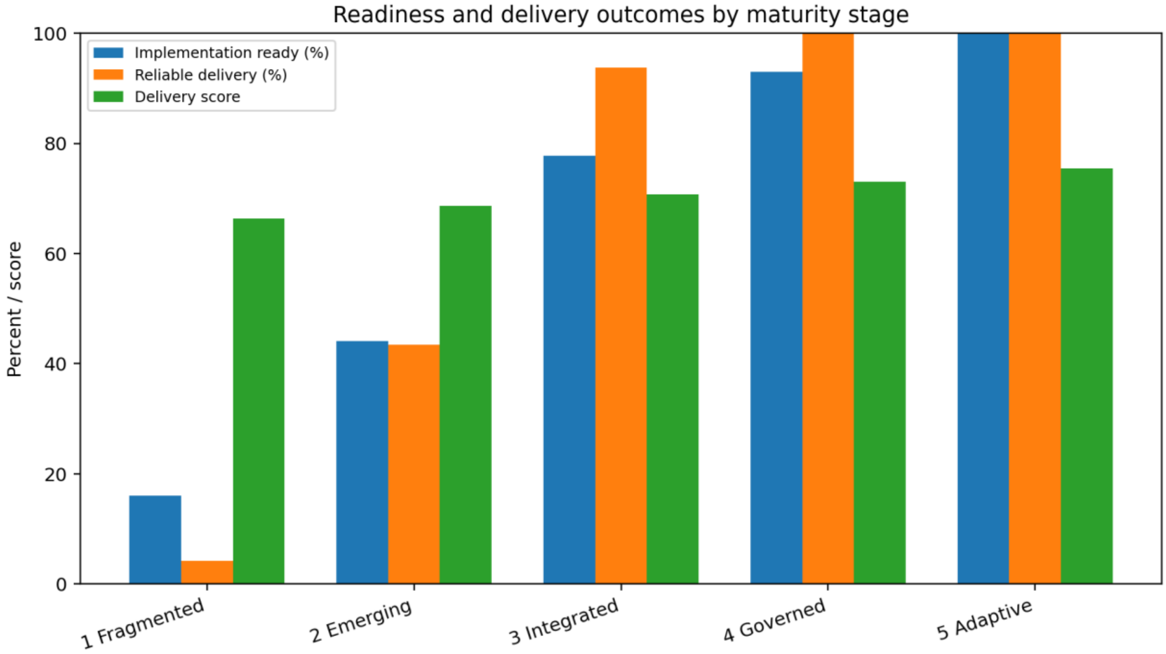

Figure 6. Readiness and delivery outcomes by maturity stage

Both outcomes increase monotonically across stages, supporting the pre-pilot staging logic.

Implementation readiness increased from 16.1% in Fragmented firms to 93.0% in Governed firms; reliable delivery rose from 4.2% to 100%. The small Adaptive group also reached 100%, but its sample size is insufficient for a stable estimate. A field study should use stratified recruitment or oversampling to ensure adequate representation at the maturity extremes.

5.8 Weighting sensitivity

Table 13. Composite-index sensitivity to alternative weights

Scheme	Spearman rho	Mean shift	Readiness AUC	Delivery r
Equal	0.9948	0.97	0.806	0.866
Outcome-derived	1.0000	0.00	0.808	0.870
Risk-balanced	0.9970	0.74	0.809	0.864

Note: The outcome-derived index is the ranking reference. Equal and risk-balanced weights produce nearly identical rankings.

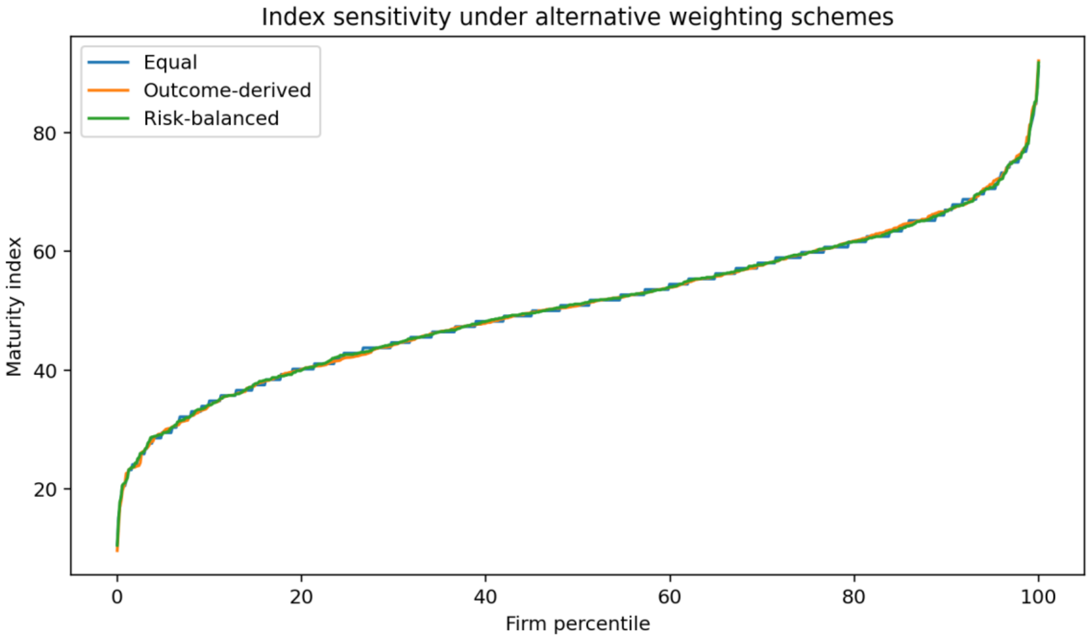

Figure 7. Sensitivity of construct weights and firm rankings

High rank correlations indicate that conclusions do not depend materially on one weighting scheme.

The minimum rank correlation across schemes was 0.995; the maximum mean absolute score shift was 0.97 points. Readiness AUC and delivery correlation were also stable. The result supports a simple provisional approach: use risk-balanced weights in practice, report all seven domains and never allow the composite to conceal a critical bottleneck.

5.9 Risk heat maps and representative cases

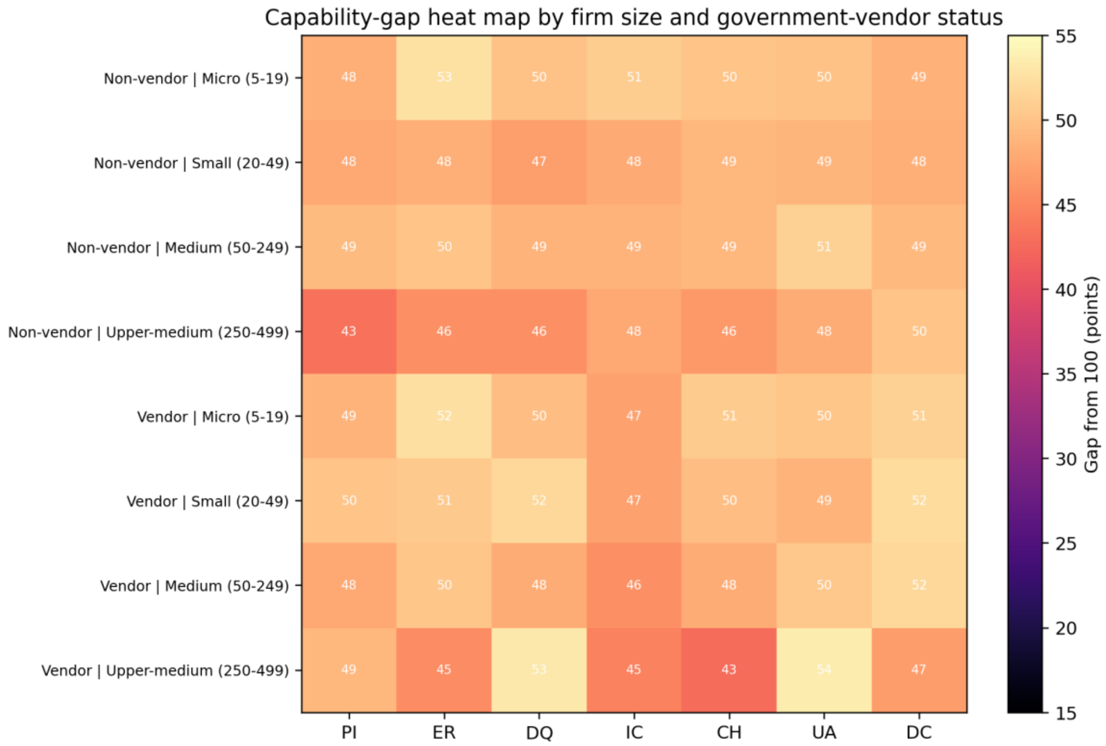

Figure 8. Average maturity gaps by firm size and government-vendor status

The heat map expresses distance from 100. It illustrates how a dashboard can direct attention toward specific capability gaps rather than only reporting an overall score.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

Table 14. Representative synthetic maturity profiles

Stage	Size	Industry	Gov.	PI	ER	DQ	IC	CH	UA	DC	Index
1 Fragmented	Medium (50-249)	Professional services	Yes	37.5	31.2	50.0	37.5	50.0	25.0	12.5	33.9
2 Emerging	Upper-medium (250-499)	Manufacturing	No	43.8	50.0	56.2	37.5	56.2	50.0	50.0	48.3
3 Integrated	Small (20-49)	Technology	No	62.5	43.8	68.8	81.2	62.5	50.0	56.2	60.8
4 Governed	Small (20-49)	Manufacturing	No	75.0	75.0	68.8	75.0	87.5	62.5	81.2	74.1
5 Adaptive	Micro (5-19)	Professional services	No	87.5	81.2	93.8	87.5	68.8	100.0	93.8	88.8

Note: Cases are nearest to each stage median. They demonstrate profile interpretation, not real organisations.

The Fragmented case illustrates a common bottleneck pattern: moderate data and cyber scores cannot compensate for very weak delivery capability, adoption and ERP readiness. The Integrated case shows a different risk: strong internal controls with weaker ERP readiness and adoption. The practical roadmap should therefore be profile-specific. A one-dimensional maturity ladder would obscure these differences.

VI. DISCUSSION

6.1 Theoretical contribution

The framework contributes to digital-transformation research by operationalising maturity as a configuration of complementary capabilities. Technology, process, data, governance, cyber, people and delivery are analytically separated but linked through outcomes. This responds to the SME literature's concern that digital transformation is context-sensitive and path-dependent rather than a universal technology sequence (Li et al., 2018; Priyono, Moin and Putri, 2020).

The results also sharpen the distinction between readiness and performance. ERP readiness, controls, data and process integration dominated the implementation-readiness equation; delivery capability, process integration and adoption dominated delivery performance. This distinction is theoretically useful because it avoids treating go-live readiness as realised value. It is consistent with information-systems success research that separates system and information quality from use and net benefits (DeLone and McLean, 2003).

The bottleneck rule extends maturity scoring beyond compensatory averages. In tightly coupled operating systems, one weak capability can constrain the whole transformation. Poor data can undermine automation; weak adoption can preserve workarounds; weak controls can destroy trust; and poor delivery governance can prevent benefits from reaching customers or contracting agencies. The minimum-domain floor therefore has conceptual and managerial justification.

6.2 Methodological contribution

The main methodological contribution is a transparent pre-pilot validation pathway. Synthetic data are often used without adequate disclosure, creating a risk that designed relationships are mistaken for empirical discovery. This paper makes the purpose, generation logic and inferential limits explicit. Reliability and factor recovery demonstrate that the computational instrument behaves as intended; they do not prove that real respondents will interpret items consistently.

The public operational benchmark adds an external scale without pretending to be an SME maturity survey. Mapping synthetic delivery ranks to an observed operational-score distribution tests dashboard behavior and enables heat-map demonstrations. The approach separates three evidence layers: construct validity, operational calibration and causal field validation. This separation parallels broader assurance principles in predictive systems: discrimination, calibration, subgroup performance, workflow utility and governance should each be tested.

Weight sensitivity addresses a recurrent maturity-model weakness. Because equal, outcome-derived and risk-balanced schemes produced almost identical rankings, practitioners can favour transparency and risk policy rather than claiming false precision from estimated weights. Field validation should nevertheless reassess weights across sectors, sizes and vendor status.

6.3 Implications for U.S. SMEs

For SMEs, the framework provides a low-burden diagnostic that can precede software selection or rescue a stalled implementation. The first managerial question becomes not which product to purchase, but which capabilities are missing. A firm with low process integration should map and simplify workflows before automating them. A firm with low data quality should establish ownership and rules before migration. A firm with low adoption should redesign roles, training and support rather than adding features.

The framework also supports financing and advisory conversations. Lenders, implementation partners, accountants and development organisations can use the same evidence vocabulary while preserving proportionality. A microbusiness does not need enterprise bureaucracy, but it does need named ownership, reliable backups, basic access controls, reconciled data and visible customer commitments. The maturity stages express increasing discipline, not increasing organisational size.

Recent SME governance research emphasises continuous controls, audit readiness and decision intelligence as resilience mechanisms (Yelduora et al., 2026; Taanisa et al., 2026a). DTM-DRF integrates those concepts with process, ERP and adoption measures, reducing the risk that digitalisation becomes a technology-only exercise.

6.4 Implications for small government vendors

Small government vendors should apply a vendor extension to the delivery-capability domain. Evidence should include contract and modification registers, milestone baselines, cost and schedule variance, deliverable acceptance, invoice-support completeness, subcontractor flow-downs, security responsibilities and corrective-action closure. The objective is not to certify compliance through a self-assessment; it is to identify evidence gaps before they become delivery or payment failures.

Government-vendor status may increase documentation and control discipline while also increasing delivery burden. The synthetic design incorporated this tension, and the risk heat map provides a practical way to examine it. A real field sample should test measurement invariance between commercial SMEs and small vendors and should avoid assuming that one benchmark applies equally across professional services, manufacturing, construction, technology and health-related services.

Cyber hygiene is especially important where vendors handle controlled data, connect to customer systems or depend on subcontractors. DTM-DRF provides a minimum readiness screen aligned to NIST's Govern-Identify-Protect-Detect-Respond-Recover logic, but specialised contractual and cybersecurity assessments remain necessary.

VII. PRACTICAL IMPLEMENTATION AND BENCHMARK DASHBOARD

7.1 Assessment workflow

Table 15. Recommended assessment-to-improvement workflow

Stage	Purpose	Core outputs
1. Governed intake	Define scope, business objective, systems, contracts, respondents and evidence boundary.	Assessment charter; stakeholder map; data inventory
2. Independent scoring	Collect ratings from process, technology, finance/control and delivery roles.	Respondent-level scores; evidence links; confidence rating

Stage	Purpose	Core outputs
3. Reconciliation workshop	Resolve material rating differences and distinguish evidence from opinion.	Agreed score; unresolved issues; owner
4. Heat-map diagnosis	Identify red domains, process/contract concentrations and maturity-stage profile.	Risk heat map; bottleneck register
5. Prioritised roadmap	Sequence foundational, enabling and optimisation actions.	90-day plan; 12-month roadmap; investment case
6. Outcome baseline	Freeze KPI definitions and baseline windows before intervention.	KPI dictionary; baseline dashboard; data-quality report
7. Review and learn	Repeat at 90 days, six months and annually; verify benefits and closure.	Trend dashboard; assurance review; revised roadmap

7.2 Dashboard architecture

Table 16. Minimum viable benchmark dashboard

Dashboard layer	Illustrative KPIs	Cadence	Accountable owner
Executive maturity	Overall index; stage; lowest domain; confidence; evidence coverage	Quarterly	Transformation sponsor
Process	Lead time; handoff defects; rework; backlog; exception ageing	Weekly/monthly	Process owner
ERP implementation	Requirements coverage; test pass rate; migration defects; cutover readiness	Weekly during project	Project manager
Data quality	Completeness; validity; duplicates; reconciliation breaks; issue ageing	Weekly/monthly	Data owner
Controls	Access exceptions; overdue reconciliations; approval overrides; repeat findings	Monthly	Controller/control owner
Cyber hygiene	MFA coverage; patch ageing; backup recovery tests; incidents; vendor reviews	Monthly/quarterly	Security owner
Adoption	Active use; process compliance; workaround rate; proficiency; support trend	Weekly/monthly	Change lead

Dashboard layer	Illustrative KPIs	Cadence	Accountable owner
Delivery	On-time milestone/delivery; invoice rejection; documentation completeness; corrective-action closure	Weekly/monthly	Delivery/contract owner

Note: All KPIs require definitions, data lineage, minimum sample thresholds and a named remediation owner.

7.3 Risk interpretation rules

- Red bottleneck: any domain below 40; immediate stabilisation action required.
- Amber gap: domain 40-54.9; defined practices exist but evidence or consistency is insufficient.
- Green capability: domain 70 or above with at least 80% evidence coverage and no overdue critical actions.
- Confidence flag: low when fewer than two informed respondents participate or evidence coverage is below 60%.
- Contradiction flag: respondent ratings differ by more than one Likert point on a critical item.
- Outcome flag: maturity improves while operational KPI performance deteriorates; investigate metric gaming, lag, scope change or external shocks.
- Reassessment flag: material ERP release, acquisition, contract change, cyber incident or process redesign triggers an interim assessment.

7.4 Ninety-day and twelve-month roadmap

Table 17. Implementation roadmap

Period	Theme	Priority actions
0-30 days	Stabilise	Confirm scope; inventory systems/data/accounts; identify critical workflows; freeze KPI definitions; remediate urgent backup/access and delivery-evidence gaps.
31-60 days	Standardise	Map future-state processes; assign owners; define master data and controls; establish issue log; design role-based training and super-user support.
61-90 days	Prove	Run controlled pilot; test migration/integration; measure adoption; verify reconciliations and invoice packs; publish baseline and post-pilot dashboard.
3-6 months	Integrate	Expand validated workflows; automate high-volume controls; close repeat findings; stabilise support; compare outcomes with baseline/control units.

Period	Theme	Priority actions
6-12 months	Govern and scale	Independent assurance; cross-site benchmarking; contract/sector templates; benefits review; recalibrate instrument; publish anonymised case study where authorised.

Management decision rule
Do not approve a major ERP expansion solely because the composite score rises. Expansion should require: no critical-domain score below 40; defined process ownership; acceptable data-migration quality; tested recovery and access controls; role proficiency; stable support; and evidence that delivery KPIs are not deteriorating.

VIII. FIELD-VALIDATION PROTOCOL

8.1 Expert review and cognitive testing

The next phase should recruit 8-12 independent experts spanning SME operations, ERP implementation, data governance, internal control, cybersecurity, organisational change and government contracting. Experts should rate item relevance, clarity and essentiality. Item-level and scale-level content-validity indices should be reported, along with Lawshe-style content-validity ratios where appropriate. Cognitive interviews with 12-20 target respondents should test interpretation, evidence availability and response burden. Items should be revised before large-scale administration.

8.2 Consenting pilot survey

Table 18. Proposed U.S. pilot validation design

Element	Specification
Target population	U.S. firms with 5-499 employees, including a purposive small-government-vendor stratum.
Sample	Minimum 400 organisations for stable factor analysis; target 600-800 to support split-sample EFA/CFA and subgroup analysis. This overall target is not sufficient by itself for the measurement-invariance testing specified in Table 19: configural, metric and scalar invariance across firm size, sector and vendor status conventionally require at least 150-200 respondents per compared group. Recruitment should therefore use a quota design (e.g., a minimum of 150 per size category and per vendor-status category) rather than free convenience sampling against the aggregate target alone, and the invariance analysis should be limited to the coarsest groupings (vendor status; broad size band) that the achieved sample can support.
Respondents	At least two knowledgeable respondents per firm where feasible: operational/technology and finance/control/delivery.
Evidence	Document references or system screenshots recorded as evidence categories, without collecting unnecessary sensitive content.
Outcomes	Implementation milestones, cycle time, data-quality exceptions, control failures, adoption, invoice rejection, on-time delivery and corrective-action closure.
Timing	Baseline, 3 months, 6 months and 12 months.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

Element	Specification
Analysis	EFA/CFA, measurement invariance, multilevel reliability, predictive validity, calibration, missing-data analysis and sector/size/vendor subgroup tests.

8.3 Case validation and causal evaluation

A nested case study should select firms from contrasting profiles rather than only contrasting total scores. Each case should document baseline process maps, evidence coverage, intervention, implementation fidelity, adoption, contextual shocks and outcomes. A matched difference-in-differences design is recommended where comparable non-implementing or later-implementing units exist. Pre-trend testing, common KPI definitions and sensitivity to matching choices are required. Interrupted time-series analysis may be preferable when a long monthly baseline is available.

8.4 Validation thresholds

Table 19. Field-validation criteria

Property	Proposed evaluation rule
Internal consistency	Alpha/omega generally 0.70-0.90; investigate redundancy above 0.95.
Convergent validity	AVE near or above 0.50 with theoretically coherent loadings.
Discriminant validity	HTMT below 0.85/0.90; test confidence intervals and cross-loadings.
Predictive validity	Cross-validated improvement over simple baselines; report calibration, not only discrimination.
Test-retest reliability	Stable scores over 2-4 weeks where no intervention occurs; target ICC of at least 0.75 for domain scores and the composite index.
Inter-rater reliability	ICC or weighted agreement adequate for organisational-level use.
Measurement invariance	Configural, metric and scalar tests across size, sector and vendor status.
Outcome sensitivity	Scores should detect credible change while resisting changes unsupported by evidence.

Note: Thresholds are interpretive guides. Decisions should consider theory, confidence intervals, sample size and consequences of misclassification.

IX. LIMITATIONS

The principal limitation is that the psychometric panel is synthetic. Designed data can test code, scoring and recoverability but cannot establish how real respondents understand items, how organisations differ, or whether maturity predicts outcomes in practice. The paper therefore uses the term pre-pilot validation and specifies a field protocol.

The DataCo extract is an international supply-chain transaction dataset rather than a U.S. SME or government-vendor sample. It contains operational indicators but no direct measures of digital maturity, ERP implementation date, training, cyber governance or contractual status. It calibrates delivery-score distributions and heat-map methods only.

The 70,000-row extract is smaller than the full 180,519-record DataCo dataset and may not preserve every distribution exactly. Results should be reproduced on the complete source and additional datasets before external benchmark publication.

The outcome-derived weights are estimated from the synthetic data-generating process and must not be presented as natural laws. Sensitivity results support provisional risk-balanced or equal weights until external coefficients are available.

Self-assessment creates risks of optimism, common-method variance and strategic rating. Multi-respondent collection, evidence coverage, independent review and outcome linkage are necessary safeguards.

The framework is broad and does not replace specialised ERP assurance, financial audit, cybersecurity assessment, legal review or government-contract compliance analysis. Sector-specific modules should be added without weakening the core measurement model.

Maturity is dynamic. A cross-sectional score can become stale after system changes, turnover, acquisition, contract shifts or cyber incidents. Repeated measurement and version control are required.

X. CONCLUSION

This study develops a practical digital-transformation maturity and delivery-readiness framework designed for U.S. SMEs and small government vendors. Its central contribution is to replace technology-adoption counting with a seven-domain capability model covering process integration, ERP readiness, data quality, internal controls, cyber hygiene, user adoption and delivery capability.

Transparent synthetic pre-pilot testing showed strong designed reliability, recovery of the intended factor structure, discriminant separation, stable weighting and coherent prediction of readiness and delivery outcomes. Public DataCo analysis demonstrated how operational performance can calibrate scores and produce cross-segment risk heat maps. These findings establish that the instrument and analytics are computationally plausible; they do not establish population prevalence or causal impact.

The practical message is direct. Digital transformation becomes operational capability only when technology is joined to governed processes, trusted data, effective controls, secure practices, competent users and reliable delivery. The framework gives firms a way to diagnose that system, sequence improvement and measure whether implementation is producing evidence-backed results. The next scholarly step is independent expert review and a preregistered, multi-site U.S. field validation.

Declarations

Funding: No external funding was received for this study.

Competing interests: The author declares no competing interests.

Ethics: The present study used public operational data and synthetic records and did not involve human participants. Future pilot surveys should obtain appropriate ethics review and informed consent.

Data availability: The DataCo SMART Supply Chain dataset is publicly available from Mendeley Data and is mirrored on Kaggle. The analysed extract, synthetic-generation specification, scoring logic and derived tables should be archived with the submitted manuscript subject to repository and licensing requirements.

Use limitation: The framework is a research and management diagnostic. It is not a certification, audit opinion, cybersecurity attestation, legal determination or guarantee of contract or implementation success.

REFERENCES

[1] 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. doi:10.1016/j.im.2007.05.004.

[2] Barney, J. (1991) 'Firm resources and sustained competitive advantage', Journal of Management, 17(1), pp. 99-120. doi:10.1177/014920639101700108.

[3] Becker, J., Knackstedt, R. and Poppelbuss, J. (2009) 'Developing maturity models for IT management: A procedure model and its application', Business & Information Systems Engineering, 1(3), pp. 213-222. doi:10.1007/s12599-009-0044-5.

[4] Bianchini, M. and Lasheras Sancho, M. (2025) SME digitalisation for competitiveness: The 2025 OECD D4SME Survey. OECD SME and Entrepreneurship Papers No. 68. Paris: OECD Publishing. doi:10.1787/197e3077-en.

[5] Bouwman, H., Nikou, S. and de Reuver, M. (2019) 'Digitalization, business models, and SMEs: How do business model innovation

practices improve performance of digitalizing SMEs?', Telecommunications Policy, 43(1), pp. 35-54. doi:10.1016/j.telpol.2017.09.007.

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

[7] Churchill, G.A. (1979) 'A paradigm for developing better measures of marketing constructs', Journal of Marketing Research, 16(1), pp. 64-73. doi:10.1177/002224377901600110.

[8] Committee of Sponsoring Organizations of the Treadway Commission (COSO) (2013) Internal Control - Integrated Framework. Durham, NC: COSO.

[9] Constante, F., Silva, F. and Pereira, A. (2019) DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS. Mendeley Data, Version 5. doi:10.17632/8gx2fvg2k6.5.

[10] Cronbach, L.J. (1951) 'Coefficient alpha and the internal structure of tests', Psychometrika, 16, pp. 297-334. doi:10.1007/BF02310555.

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

[12] 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. doi:10.1080/07421222.2003.11045748.

[13] Eliot, D. (2024) NIST Cybersecurity Framework 2.0: Small Business Quick-Start Guide. NIST SP 1300. Gaithersburg, MD: National Institute of Standards and Technology. doi:10.6028/NIST.SP.1300.

[14] Eliot, D., Marron, J. and Thorn, S. (2026) Small Business Cybersecurity: Non-Employer Firms. NIST CSWP 50, Initial Public Draft. Gaithersburg, MD: National Institute of Standards and Technology. doi:10.6028/NIST.CSWP.50.ipd.

[15] 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. doi:10.1016/j.jbusres.2020.03.004.

[16] Fornell, C. and Larcker, D.F. (1981) 'Evaluating structural equation models with unobservable variables and measurement error', Journal of Marketing Research, 18(1), pp. 39-50. doi:10.1177/002224378101800104.

[17] Garzoni, A., De Turi, I., Secundo, G. and Del Vecchio, P. (2020) 'Fostering digital transformation of SMEs: A four levels approach', Management Decision, 58(8), pp. 1543-1562. doi:10.1108/MD-07-2019-0939.

[18] 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. doi:10.2307/25148695.

[19] Hair, J.F., Risher, J.J., Sarstedt, M. and Ringle, C.M. (2019) 'When to use and how to report the results of PLS-SEM', European Business Review, 31(1), pp. 2-24. doi:10.1108/EBR-11-2018-0203.

[20] Henseler, J., Ringle, C.M. and Sarstedt, M. (2015) 'A new criterion for assessing discriminant validity in variance-based structural equation modeling', Journal of the Academy of Marketing Science, 43, pp. 115-135. doi:10.1007/s11747-014-0403-8.

[21] Hinkin, T.R. (1998) 'A brief tutorial on the development of measures for use in survey questionnaires', Organizational Research Methods, 1(1), pp. 104-121. doi:10.1177/109442819800100106.

[22] Ifinedo, P. (2011) 'Internet/e-business technologies acceptance in Canada's SMEs: An exploratory investigation', Internet Research, 21(3), pp. 255-281. doi:10.1108/10662241111139309.

[23] Kaiser, H.F. (1974) 'An index of factorial simplicity', Psychometrika, 39, pp. 31-36. doi:10.1007/BF02291575.

[24] Li, L., Su, F., Zhang, W. and Mao, J.Y. (2018) 'Digital transformation by SME entrepreneurs: A capability perspective', Information Systems Journal, 28(6), pp. 1129-1157. doi:10.1111/isj.12153.

[25] MacKenzie, S.B., Podsakoff, P.M. and Podsakoff, N.P. (2011) 'Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques', MIS Quarterly, 35(2), pp. 293-334. doi:10.2307/23044045.

[26] Matarazzo, M., Penco, L., Profumo, G. and Quaglia, R. (2021) 'Digital transformation and customer value creation in Made in Italy SMEs: A dynamic capabilities perspective', Journal of Business Research, 123, pp. 642-656. doi:10.1016/j.jbusres.2020.10.033.

[27] Mettler, T. (2011) 'Maturity assessment models: A design science research approach', International Journal of Society Systems Science, 3(1/2), pp. 81-98. doi:10.1504/IJSSS.2011.038934.

[28] Oliveira, T. and Martins, M.F. (2011) 'Literature review of information technology adoption models at firm level', Electronic Commerce Research and Applications, 10(3), pp. 110-121. doi:10.1016/j.elerap.2010.12.001.

[29] Parasuraman, A. (2000) 'Technology Readiness Index (TRI): A multiple-item scale to measure readiness to embrace new technologies', Journal of Service Research, 2(4), pp. 307-320. doi:10.1177/109467050024001.

[30] Pascoe, C., Quinn, S. and Scarfone, K. (2024) The NIST Cybersecurity Framework (CSF) 2.0. NIST CSWP 29. Gaithersburg, MD: National Institute of Standards and Technology. doi:10.6028/NIST.CSWP.29.

[31] Poppelbuss, J. and Roglinger, M. (2011) 'What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management', Proceedings of the 19th European Conference on Information Systems, Helsinki.

[32] Priyono, A., Moin, A. and Putri, V.N.A.O. (2020) 'Identifying digital transformation paths in the business model of SMEs during the COVID-19 pandemic', Journal of Open Innovation: Technology, Market, and Complexity, 6(4), article 104. doi:10.3390/joitmc6040104.

[33] 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. doi:10.1016/j.accinf.2014.01.002.

[34] Sambamurthy, V., Bharadwaj, A. and Grover, V. (2003) 'Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms', MIS Quarterly, 27(2), pp. 237-263. doi:10.2307/30036530.

[35] Taanisa, T., Mukwata, N.A., Sydney, J., Ganyani, L., Maturure, R.N., Chingezi, E., Yelduora, P.G. and Mupa, M.N. (2026a) 'AI-enabled audit analytics for SME financial reporting and anomaly detection: A risk-based framework for early irregularity identification and control strengthening', World Journal of Advanced Research and Reviews, 30(3), pp. 1113-1126. doi:10.30574/wjarr.2026.30.3.1596.

[36] Teece, D.J. (2007) 'Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance', Strategic Management Journal, 28(13), pp. 1319-1350. doi:10.1002/smj.640.

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

[38] 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. doi:10.1016/S0377-2217(02)00547-7.

[39] 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. doi:10.2307/30036540.

[40] Verhoef, P.C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J.Q., Fabian, N. and Haenlein, M. (2021) 'Digital transformation: A multidisciplinary reflection and research agenda', Journal of Business Research, 122, pp. 889-901. doi:10.1016/j.jbusres.2019.09.022.

[41] Vial, G. (2019) 'Understanding digital transformation: A review and a research agenda', Journal of Strategic Information Systems, 28(2), pp. 118-144. doi:10.1016/j.jsis.2019.01.003.

[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. doi:10.1080/07421222.1996.11518099.

[43] Warner, K.S.R. and Wager, M. (2019) 'Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal', Long Range Planning, 52(3), pp. 326-349. doi:10.1016/j.lrp.2018.12.001.

[44] 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. doi:10.1108/02635570210414668.
[45] Yelduora, P.G., Nhemachena, T.L., Chingezi, E. and Mupa, M.N. (2026) 'Data-driven internal controls and budget governance for small businesses and nonprofit institutions: A practical model for audit readiness and financial resilience', Iconic Research and Engineering Journals, 10(1). doi:10.64388/IREV10I1-1719493.

Appendix A. Full instrument and evidence guide

Table 20. Item-level scoring and evidence guide

Item	Statement	Minimum evidence example	Anchor
PI1	End-to-end workflows are documented and owned across functions.	Approved process map/process owner	1=absent; 3=defined/inconsistent; 5=embedded/measured
PI2	Handoffs and exception routes are standardised.	SOP/handoff matrix	1=absent; 3=defined/inconsistent; 5=embedded/measured
PI3	Core systems exchange data without repeated manual re-entry.	Interface or integration log	1=absent; 3=defined/inconsistent; 5=embedded/measured
PI4	Rework, queue time and process exceptions are measured and reviewed.	Rework/exception dashboard	1=absent; 3=defined/inconsistent; 5=embedded/measured
ER1	A quantified business case, scope and process ownership are approved.	Business case/charter	1=absent; 3=defined/inconsistent; 5=embedded/measured
ER2	Fit-gap requirements and vendor-neutral acceptance criteria are documented.	Requirements and acceptance criteria	1=absent; 3=defined/inconsistent; 5=embedded/measured
ER3	Resources, change impacts, risks and decision rights are established.	Resource/risk/change plan	1=absent; 3=defined/inconsistent; 5=embedded/measured
ER4	Data migration, testing, cutover and stabilisation plans are ready and rehearsed.	Migration, test and cutover pack	1=absent; 3=defined/inconsistent; 5=embedded/measured
DQ1	Master-data domains have named owners and definitions.	Data-owner register/dictionary	1=absent; 3=defined/inconsistent; 5=embedded/measured
DQ2	Validation rules address duplicates, invalid values and missing critical fields.	Validation and duplicate rules	1=absent; 3=defined/inconsistent; 5=embedded/measured

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

Item	Statement	Minimum evidence example	Anchor
DQ3	Completeness, accuracy, consistency and timeliness are monitored.	Quality scorecard	1=absent; 3=defined/inconsistent; 5=embedded/measured
DQ4	Source-to-report lineage and reconciliations are documented.	Lineage/reconciliation evidence	1=absent; 3=defined/inconsistent; 5=embedded/measured
IC1	Role-based access and segregation-of-duties risks are reviewed.	Access and SoD review	1=absent; 3=defined/inconsistent; 5=embedded/measured
IC2	Approvals, thresholds and supporting evidence are enforced.	Approval matrix/audit trail	1=absent; 3=defined/inconsistent; 5=embedded/measured
IC3	Reconciliations and exception monitoring occur at defined frequencies.	Reconciliation/exception log	1=absent; 3=defined/inconsistent; 5=embedded/measured
IC4	Control failures have owners, due dates, root causes and verified closure.	Root-cause and closure evidence	1=absent; 3=defined/inconsistent; 5=embedded/measured
CH1	Critical assets, accounts and data are inventoried; strong authentication is used.	Asset/account inventory and MFA report	1=absent; 3=defined/inconsistent; 5=embedded/measured
CH2	Backups are protected and recovery is tested.	Backup/recovery test	1=absent; 3=defined/inconsistent; 5=embedded/measured
CH3	Patching, vulnerability handling and incident escalation are defined.	Patch/vulnerability/incident procedure	1=absent; 3=defined/inconsistent; 5=embedded/measured
CH4	Third-party risk and workforce security awareness are reviewed.	Vendor-risk review and awareness record	1=absent; 3=defined/inconsistent; 5=embedded/measured
UA1	Training is role-based and verifies practical proficiency.	Role curriculum/proficiency result	1=absent; 3=defined/inconsistent; 5=embedded/measured
UA2	Super users and support routes are available after go-live.	Super-user/support roster	1=absent; 3=defined/inconsistent; 5=embedded/measured
UA3	Actual system use, workarounds and process compliance are measured.	Usage/workaround/compliance metric	1=absent; 3=defined/inconsistent; 5=embedded/measured

Item	Statement	Minimum evidence example	Anchor
UA4	User feedback, support trends and lessons learned drive improvement.	Feedback and improvement log	1=absent; 3=defined/inconsistent; 5=embedded/measured
DC1	Schedule, cost, milestones and delivery commitments are actively controlled.	Schedule/cost/milestone baseline	1=absent; 3=defined/inconsistent; 5=embedded/measured
DC2	Order-to-cash and invoice documentation are complete and timely.	Order-to-cash/invoice checklist	1=absent; 3=defined/inconsistent; 5=embedded/measured
DC3	Quality, delivery and corrective-action performance are monitored.	Quality/delivery/corrective-action dashboard	1=absent; 3=defined/inconsistent; 5=embedded/measured
DC4	Subcontractor, contract and evidence obligations are traceable and current.	Subcontract/contract evidence register	1=absent; 3=defined/inconsistent; 5=embedded/measured

Note: Assessors should record evidence date, owner and confidence. A rating of 4 or 5 should not be assigned without corroborating evidence.

Appendix B. Scoring formulas and dashboard rules

Table 21. Core formulas

Metric	Formula	Output
Domain score	((mean of four item ratings - 1) / 4) x 100	0-100
Composite index	Sum(domain score x domain weight)	0-100
Evidence coverage	Supported items / applicable items x 100	0-100%
Bottleneck gap	100 - lowest domain score	0-100
Contradiction flag	Absolute inter-rater difference >1 on a critical item	Yes/No
Reliable delivery	Delivery KPI meets pre-defined, evidence-backed threshold	Yes/No
Change score	Current index - baseline index	Points with confidence interval
Benefits realisation	Verified monetary/non-monetary benefit / approved target	Percent

Recommended provisional risk-balanced weights are PI 18%, ER 16%, DQ 16%, IC 14%, CH 12%, UA 12% and DC 12%. Report the unweighted domain scores beside the composite. Apply the maturity-stage thresholds only after bottleneck, evidence-coverage and confidence rules are checked.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1722514

Appendix C. Government-vendor extension

Table 22. Optional small-government-vendor extension

Item	Assessment statement
GV1	Contract, task-order and modification obligations are maintained in a current register.
GV2	Milestone, schedule and cost baselines are approved and variance is reviewed.
GV3	Deliverable acceptance criteria and evidence are traceable to each commitment.
GV4	Invoice packs are complete, timely and linked to accepted work and authorised costs.
GV5	Subcontractor obligations, performance and required flow-downs are monitored.
GV6	Security, privacy and data-handling responsibilities are assigned and evidenced.
GV7	Corrective actions have owners, root causes, due dates and verified closure.
GV8	Management reviews contract risks, customer communication and forecast-at-completion.

Note: These items may form a separate vendor-readiness module. They should not be used as a legal or contractual compliance certification.

Appendix D. Reproducibility and analysis specification

12. Obtain the public DataCo source or document the exact analytical extract and checksum.
13. Parse order dates; derive severe delay, hard workflow exceptions, negative-profit exposure and critical-field completeness.
14. Aggregate market x department x segment x year-quarter units and retain at least 60 transaction lines per unit.
15. Compute the six-component operational delivery score using the weights in Table 7.
16. Generate or import item-level assessment records; retain raw item values and respondent identifiers separate from organisational aggregates.
17. Compute domain scores, evidence coverage, confidence and alternative composite indices.
18. Estimate alpha, composite reliability, AVE, KMO, Bartlett test, factor structure and HTMT.
19. Use split-sample or cross-validated predictive models and report calibration alongside discrimination.
20. Conduct sensitivity analysis for weights, thresholds, missing-data rules and subgroup definitions.
21. Archive code, data dictionary, instrument version, random seed, analytic decisions and limitations.

Submission note

Before journal submission, the synthetic pre-pilot results should be accompanied by the analysis code and clearly labelled in the title, abstract, methods, tables and data statement. After field data are collected, the synthetic section should be retained as instrument-development evidence or moved to supplementary materials rather than blended with observed results.

Data-source links

Mendeley Data: DataCo SMART Supply Chain dataset

Kaggle mirror: DataCo SMART Supply Chain for Big Data Analysis

NIST CSF 2.0: Cybersecurity Framework 2.0