The Smart Manufacturing Organizational Capability Model: a framework for assessing change-execution and sustainment capability across firm sizes

Suraj Sriram

Independent Researcher, Atlanta, Georgia, USA

ORCID: 0009-0002-9064-3853 · Correspondence: suraj28@gmail.com

Status of this preprint. This paper presents a designed artifact and its full specification prior to empirical validation. A two-stage validation study, comprising expert interviews followed by a stratified practitioner survey, is in preparation and will be reported separately. No empirical findings are claimed here. The complete assessment criteria are included so that the framework can be examined, applied and criticized in its current form.

Abstract

Industry survey evidence indicates that under a third of manufacturers are scaling smart factory initiatives across the enterprise, and that a large majority cannot scale digital innovation across their manufacturing networks [1,2]. Two failure modes underline these figures. Initiatives may never escape pilot stage. Alternatively they scale successfully and then lose realized value in the months after deployment. The second is less studied than the first, although it consumes investment already committed.

Both failures are organizational rather than technical, and the maturity literature has responded by adding organizational dimensions to instruments that were previously technology focused. This paper argues that the organizational category so constructed contains two constructs rather than one. Steady-state operational capability concerns how well an organization runs what it already has. Change-execution capability concerns whether it can deploy something new, retain the ability to operate it, absorb the disruption of adopting it, and confirm that projected value survived. Existing instruments assess the first. None assess the second.

The Smart Manufacturing Organizational Capability Model (SM-OCM) is a five-dimension, five-stage framework addressing change-execution and sustainment capability, specified to hold across the size distribution of the manufacturing base. Constructs are stated independently of organization scale. The evidence demonstrating a construct at a given stage is banded by firm size, so that a forty-person manufacturer and a seventy-site enterprise are assessed on the same capabilities against conditions appropriate to their scale. Sixty capability evidence criteria are specified in full.

The contribution is the distinction between the two capability types. The instrument operationalizes it so that the distinction becomes assessable and testable.

Keywords: smart manufacturing; maturity model; organizational capability; digital transformation;

1. Introduction

Manufacturing enterprises have invested heavily in smart manufacturing technologies over the past decade, and the returns have been uneven in a well-documented pattern. Industry survey evidence indicates that under a third of manufacturers are successfully scaling smart factory initiatives across the enterprise, and international research on manufacturing networks finds a large majority unable to scale digital innovations beyond initial sites. The resulting condition has been characterized as pilot purgatory: demonstrations succeed, replication does not, and investment accumulates without commensurate return.

The explanation offered is consistently organizational. Surveys and practitioner accounts attribute scaling failure to workforce capability, governance, change management and absent ownership rather than to technology that does not work. Practitioner training material in this field commonly apportions digital transformation failure to people, process and technology in roughly a 63:30:7 ratio. Whatever the precision of such figures, the direction is not in dispute. What remains missing is an instrument for acting on the diagnosis. Existing assessments measure steady-state operational maturity rather than an organization's capability to execute and sustain change within it.

A second failure mode has received considerably less attention. Beyond initiatives that never escape pilot stage are initiatives that scale successfully and then lose value in the months after deployment. The system functions as designed, but usage decays. Operators revert to prior methods. A dashboard nobody consults costs attention and returns nothing. The business case is rarely invalidated formally. It stops being true after the deployment team has moved on, and often nobody checks. This failure is distinct from pilot failure and is more costly, because the investment has already been made in full.

Both failure modes trace to the same underlying construct: the organization's capability to deploy, adopt, sustain and monetize technology. This is distinct from the state of the technology itself, and from the steady-state condition of the operation around it.

1.1 The scale problem

Existing maturity models are mostly developed from and calibrated to large-enterprise contexts. The SME-focused maturity literature identifies this directly [5,6]. It matters because the manufacturing base is not large. Around ninety-eight percent of United States manufacturers employ fewer than five hundred people, and public technology extension programs serving this population increasingly focus on firms below twenty employees. A maturity model whose indicators assume a Center of Excellence, portfolio governance or a dedicated specialist team cannot be used by most of the firms that need it.

This paper specifies capability constructs at a level of abstraction that holds independent of organization size, with evidence criteria banded by scale. The constructs are common to all sizes. What counts as observable evidence of a given stage differs between a forty-person single-plant manufacturer and a seventy-site global enterprise, while the capability being assessed does not.

Research questions

(1) Which organizational capability dimensions are implicated in whether smart manufacturing technology investments scale across an organization and retain realized value after deployment?

(2) Can these dimensions be specified as a maturity framework with stages that discriminate meaningfully and advancement criteria that are observable rather than subjective?

(3) Can such a framework be specified to hold across smart manufacturing technology domains and across organization scales?

This paper addresses the first two questions by construction and specification. The third is an empirical question which the framework is designed to make testable and which the planned validation described in Section 6 is intended to answer.

2. Related Work and the Organizational Capability Gap

2.1 Review scope

This review is cumulative rather than exhaustive. Vance and colleagues [4] examined nineteen peer-reviewed models and five consulting-firm models, deriving ten dimension categories against which they compared them. Replicating that work would add nothing. This review takes their classification as its baseline and asks whether the derived categories accommodate the constructs proposed here.

2.2 What existing instruments assess

The classification groups maturity model dimensions into ten categories: information technology, cyber-physical systems and data; strategy and organization; supply chain and logistics; products and services; culture and employees; technology and capabilities; customer and market; cybersecurity and risk; leadership and management; and governance and compliance [4].

Two features of that set matter here. It is comprehensive with respect to what existing models assess, having been derived inductively from twenty-four instruments rather than imposed on them. And every category describes a property of the organization or its technology in a steady state. None describes the capability to execute and sustain a change to that state.

The same review reports that the models examined were weak in cybersecurity, leadership and governance, and that many have not been validated or do not provide a usable tool for application. Peer-reviewed models lacked industry awareness and were difficult to apply. Consulting-firm models were more applicable but lacked published detail and validation. The authors conclude by inviting researchers to determine, on that basis, whether an existing model applies or whether a new one is required [4].

2.3 Relationship to SME-focused maturity research

A separate research stream addresses the applicability of smart manufacturing and Industry 4.0 maturity models to small and medium-sized enterprises, and proposes SME-specific instruments [5,6]. That work correctly identifies that large-enterprise-derived models transfer poorly to resource-constrained contexts. It resolves the problem by building separate models.

This paper takes a different position. Separate models by size prevent comparison across a supply network in which large manufacturers and their small suppliers must progress together. If the underlying capability constructs are organizational rather than resource-dependent, one framework with scale-banded evidence criteria is preferable to two with incommensurable constructs. The framework is designed to make this testable rather than to assert it.

2.4 The gap

The maturity literature does not ignore organizational factors. Schumacher and colleagues [7] added strategy, leadership, governance, culture and people dimensions to a nine-dimension Industry 4.0 instrument for this reason, and capability models derived from the technology-organization-environment framework place organization alongside technology as a determinant of transformation capability [8]. This turn in the literature is more than a decade old.

The boundary that remains sits inside the organizational category rather than outside it. Existing organizational dimensions assess attributes of the operation in its steady state: whether a digital strategy exists, whether leadership is committed, whether the culture supports improvement, whether the workforce is skilled for its current work. They do not assess whether the organization can execute and sustain a particular technology change. Whether a strategy exists is a different question from whether initiatives are sequenced against measured absorption limits.

Mapping the constructs proposed in this paper against the ten categories makes the boundary explicit.

Proposed construct	Nearest category in the published ten	Status
D1 Capability Ownership and Continuity	Culture and employees; leadership and management	Not covered. Both categories describe the workforce and leadership of the operation as it stands. Neither addresses whether capability to operate a deployed system persists
D2 Investment Governance and Sequencing	Governance and compliance; strategy and organization	Partially covered, and identified in the review as among the weakest areas across the models examined [4]. Sequencing under capacity constraint is absent
D3 Change Enablement and Workforce Readiness	Culture and employees	Not covered. The category concerns workforce skills and culture in steady state. It does not address capacity to absorb a change, or willingness to act on system output
D4 Operational Continuity of Deployed Systems	IT, CPS and data; cybersecurity and risk	Partially covered. Security is addressed. Availability governance, tested recovery and uptime accountability are not
D5 Value Realization and Sustainment	None	Not covered. Consulting instruments were within the review's scope, so this absence does not follow from restricting attention to academic models

Table 1. Proposed constructs mapped against the ten dimension categories derived in the published review [4].

Three of the five constructs have no counterpart. The two partial overlaps concern governance and infrastructure, which straddle both the steady state and the execution of change, so their position is consistent with the distinction rather than an exception to it.

Stated directly: existing smart manufacturing maturity models assess the state of technology and the steady-state organizational condition of the operation, including its strategy, leadership, governance, culture and workforce skills. None assess change execution and sustainment capability. None are specified to hold across the full size distribution of the manufacturing base.

This is narrower than a claim that the literature neglects organizational factors, and it is the claim the evidence supports.

2.5 Relationship to the author's prior practitioner publications

Two prior practitioner publications by the author bear on this work. The first presents a single-enterprise case study deriving a three-pillar scaling framework of people, process and data foundations, mapped to an existing smart manufacturing roadmap framework. The second argues that post-deployment value erosion arises from change fatigue, unmanaged absorption capacity and uncalibrated operator trust in system recommendations.

Neither constitutes a maturity framework. The first is a descriptive account of one enterprise's trajectory, validated by mapping rather than by independent assessment; the second asserts phenomena without operationalizing them. The present work differs in artifact class, method and intended evidence base: it specifies discriminated capability constructs with observable advancement criteria, is developed under an explicit design-science protocol, and is designed for validation across multiple independent organizations spanning size bands.

The three-pillar formulation is superseded here. People separates into D1 and D3, distinguishing the capability to staff a program from the capability to sustain workforce adoption — constructs with independent failure modes that the earlier formulation conflated. Process separates into D2, D4 and D5, distinguishing investment decision-making, operational sustainment and value accounting. Data foundations is excluded by scope, for the reason given in Section 4.2.

3. Design Approach

The framework is constructed as an artifact and then evaluated, following Design Science Research principles [9,10]. Grounded theory would misdescribe the process, since the construct set precedes data collection. Design Science treats artifact construction as the contribution and supplies an established build-and-evaluate cycle. Guidance on maturity model development within that tradition informed the procedure followed here [11,12].

Two design commitments follow from the review's findings about the state of the field [4]. Many existing models have not been validated, and those that are validated are frequently not applicable: peer-reviewed instruments lack usable assessment tooling, and consulting instruments lack published detail. This design responds to both. Empirical validation is a principal reason the work exists in this form rather than an optional strengthening step. Applicability is built in through size-banded evidence criteria and advancement conditions requiring demonstrated instances, so that an assessor can apply the framework without proprietary tooling. The criteria are published here in full for the same reason.

3.1 Practitioner evidence as design input

Construction draws on longitudinal evidence from a multi-year enterprise digital manufacturing transformation spanning more than seventy manufacturing sites, over three hundred system go-lives and more than five thousand users, together with prior multi-site deployment programs at three other manufacturing enterprises. This evidence informs artifact construction. It does not validate it. The separation is maintained throughout, because conflating construction evidence with validation evidence is the principal validity risk in practitioner-originated framework research.

The construction evidence base is predominantly large-enterprise. Any claim to scale generality therefore rests on the size-stratified validation described in Section 6, and is falsifiable by it. If small and mid-sized respondents assess the constructs as irrelevant or the banded criteria as unusable, the claim fails and will be withdrawn.

3.2 Measurement model

The framework is specified as formative rather than reflective [13,14], and the distinction governs what counts as evidence of quality. Each dimension is constituted by its evidence criteria rather than manifested by them. An organization does not possess capability ownership which then causes a named backup to exist. The existence of a named backup, protected time and written procedure together constitute the capability at that stage. Criteria are therefore not expected to be interchangeable, to covary, or to load on a common factor.

Internal consistency reliability is consequently not an appropriate measure, and reporting it on formative indicators would be a category error [14]. Criterion quality is assessed instead through indicator relevance ratings, redundancy analysis within dimensions, and a nomological check against outcome measures. This specification narrows the validation evidence available. That is a cost of the design rather than an exemption from scrutiny.

The banded design raises a measurement invariance question. If indicators differ by size band, is the same construct being measured? Under reflective specification this would require invariance testing. Under formative specification it is the expected case. The construct definition is held invariant by design and stated identically across bands, while the indicators constituting it vary with organizational context. What must be defended is that the banded indicators are each valid constitutive elements of the same construct, and this is what expert validation is designed to test.

4. The Framework

4.1 Design principle: scale-independent constructs, scale-contingent evidence

The framework separates two things that maturity models conventionally conflate. A capability construct describes what organizational property is being assessed. An evidence criterion describes what observable fact demonstrates that the property is present at a given stage. The first is held constant across organization size; the second is banded.

D1 assesses whether capability is held internally and persists beyond an individual's departure. In a large enterprise the evidence at the proactive stage is an operating Center of Excellence with a competency model and succession planning. In a sixty-person manufacturer it is a named individual with protected time, a documented backup, and written procedures sufficient for a successor to operate the system. The capability is the same in both cases. The evidence is not. Requiring the large-enterprise evidence of a small manufacturer measures headcount rather than capability.

4.2 Construct derivation rule

Design Science requires that artifact construction be traceable rather than asserted. A candidate construct was admitted if and only if all three of the following conditions held.

(4) Its absence is an observed cause of post-deployment failure in the practitioner evidence base or in the documented industry failure literature.
(5) It is assessable from observable organizational facts rather than from attitude or stated intent.
(6) It is not already covered by instruments assessing technology state or steady-state operational condition.

Candidates were rejected under this rule. The most substantial was data foundations, meaning the quality, availability and contextualization of manufacturing data, which satisfies the first two conditions comfortably and was present in the author's prior three-pillar formulation. It fails the third: data architecture maturity is assessed directly by ISA-95-aligned instruments, by interoperability specifications, and by the data-driven processes dimension of the practitioner framework examined in Section 2.4. Including it would duplicate existing coverage and breach the boundary on which this framework's contribution rests.

Two further candidates were rejected under the second condition: executive sponsorship and organizational culture. Both are cited frequently as determinants of transformation success, and both fail the observability test — in practice they are assessed through attitudinal self-report, which this framework deliberately avoids. Where they exert influence they do so through mechanisms the framework does capture: sponsorship through investment governance, culture through whether change is absorbed and superseded tools are actually retired.

One admission requires explanation against the second condition. Trust calibration concerns an attitudinal phenomenon, yet culture and sponsorship were rejected as attitudinal. The rule applies to how a construct is assessed rather than to the nature of the phenomenon beneath it. Trust calibration is admitted because a behavioural indicator exists and is used: the criterion asks whether the organization knows if operators act on system recommendations and has investigated where they do not. It does not ask whether operators report trusting the system.

Recording these rejections is the evidence that the rule was applied rather than reverse-engineered from a construct set already chosen.

4.3 Dimension constructs

ID	Dimension	Scale-independent construct	Failure mode addressed
D1	Capability Ownership & Continuity	Whether the capability to operate and adapt deployed technology is held within the organization and persists independently of any individual or external party, and when accountability for post-deployment outcomes is established	Capability resides externally or in one person; knowledge does not persist; outcome ownership is handed over after go-live rather than settled before it
D2	Investment Governance & Sequencing	Whether technology investment decisions are made against stated criteria, sequenced deliberately rather than concurrently, and standardized across the organization's operating units	Initiatives proliferate without a mechanism converting them to standard practice; concurrent launches exceed what the organization can absorb
D3	Change Enablement & Workforce Readiness	Whether the workforce is prepared, enabled and measured in its capacity to adopt and sustain a technology change, distinct from being equipped for steady-state operations — comprising readiness (competence to use what is newly deployed), change absorption capacity (remaining organizational capacity to accept further change), and trust calibration (willingness to act on system output)	Technology deployed to a workforce with no remaining absorption capacity; usage decays after go-live; operators understand the tool but withhold trust in its recommendations
D4	Operational Continuity of Deployed Systems	Whether availability, performance and recoverability of deployed digital assets are governed rather than left to chance or to the vendor	Deployed systems degrade; confidence erodes; users revert to prior methods
D5	Value Realization & Sustainment	Whether projected value is baselined, tracked after deployment rather than at it, and converted into subsequent investment decisions	Value confirmed at go-live rather than where it holds or leaks; sustained return undemonstrable; funding withdrawn at the next cycle

Table 4. SM-OCM dimension constructs, stated independently of organization scale.

Two sub-constructs within D3 borrow from adjacent literatures and are distinguished from them here. Change absorption capacity denotes an organization's remaining capacity to accept further change. It is not absorptive capacity in the sense of Cohen and Levinthal [15], which concerns the ability to recognize, assimilate and apply external knowledge. These are unrelated constructs that share a word. The term used here is anchored instead to the change fatigue and change saturation literature [16]. Trust calibration denotes the alignment between an operator's willingness to act on system output and that system's actual reliability. It is taken from the trust-in-automation literature [17] and applied to manufacturing system recommendations rather than coined here.

4.4 Maturity stages

Stage names are adopted from vocabulary already in practitioner use: reactive, aware, collaborative, proactive, adaptive. The stage progression is not this framework's contribution. The capability constructs and their scale-banded evidence criteria are. Adopting established vocabulary allows a practitioner assessing an organization on more than one instrument to read the results together without translation.

Stage	Designation	Organizational characteristic (scale-independent)
S1	Reactive	Capability is external or absent; decisions are ad hoc; outcomes are unowned; value is asserted rather than measured
S2	Aware	Capability is recognized but informal and person-dependent; intentions are stated without mechanism; adoption is assumed
S3	Collaborative	Capability is defined and documented; decisions follow stated criteria; readiness and realization are assessed at defined points
S4	Proactive	Capability is held internally and governed; change absorption capacity is a managed constraint; realization is measured after deployment
S5	Adaptive	Capability generates improvement and is transferable; sequencing optimizes against measured absorption; value decay is detected and arrested

Table 5. SM-OCM maturity stages.

4.5 Advancement rules and the demonstrated-instance principle

A maturity framework without explicit advancement conditions is a descriptive taxonomy rather than a diagnostic instrument. For each dimension, advancing from one stage to the next requires satisfaction of all stated evidence criteria for the organization's size band. These are termed capability evidence criteria rather than transition criteria, to avoid collision with state transition matrix terminology already in practitioner use for a different artifact.

Organizations are assessed against the band corresponding to employee count at the operating unit holding technology decision authority, not corporate headcount. A three-hundred-person division with independent decision rights is assessed as mid-market.

Criteria require demonstrated instances rather than documented intentions: a backup who has operated the system rather than one who has been designated, a rejection actually made against stated criteria rather than criteria that merely exist, a deployment actually deferred, a review that changed a decision. Any criteria-based instrument can be satisfied on paper. An organization treating this framework as a compliance exercise should find the criteria hard to satisfy without the underlying behavior.

The complete specification, covering twenty stage advancements across three size bands and sixty criteria statements, appears in Appendix A.

5. Intended Use

The framework's first use is diagnostic, and it produces a profile rather than a score. Composite scoring is not provided. A single number averages away the one dimension that is usually the binding constraint, and it invites optimisation of the score rather than the capability. An organization reading its profile should attend to its lowest dimension first, since the dimensions are not substitutable. Strong governance does not compensate for capability that leaves with a departing individual, and neither compensates for a workforce with no remaining capacity to absorb another change.

What follows from a low rating differs by dimension, and the practical value of separating the constructs lies in this. A low D1 rating indicates that the organization should not proceed with further deployment until ownership and continuity are established, because each additional system deepens a dependency it cannot sustain. A low D3 rating indicates a sequencing problem rather than a training gap, and the appropriate response is to defer rather than to train harder. A low D5 rating is the most overlooked, because it is invisible at go-live. The organization is deploying successfully and cannot demonstrate that anything deployed earlier is still delivering. The intervention differs in each case, and an instrument reporting a single maturity number cannot distinguish between them.

The second use is in sequencing. The conventional roadmapping question is which initiatives to pursue and in what order, answered by business value and technical dependency. The framework adds a third input: what the organization can absorb. An initiative correctly prioritised on value and correctly sequenced on dependency will still fail if it lands on a workforce that has not recovered from the last one. Assessing change absorption capacity before scheduling turns a constraint usually discovered afterwards, as decayed adoption, into one that can be planned around.

The third use follows from D5. Value confirmed at go-live is not value realized. Organizations that assess realization at a defined interval after deployment, and act on what they find, retain value that otherwise erodes without any decision being taken to abandon it. The framework makes the timing of that assessment an explicit maturity property rather than an unstated assumption.

For advisory and extension organizations, the scale-banded design addresses a specific problem. An instrument calibrated to large enterprises does not discriminate at small scale, while an instrument built only for small firms cannot be compared against assessments of the larger organizations those firms supply. Constructs are held constant so that results are comparable across the size distribution; evidence criteria are banded so that a forty-person manufacturer is assessed on what capability looks like at forty people.

The supply-network consequence is the more significant one. Where a large manufacturer and its smaller suppliers are progressing together, as data exchange and traceability obligations increasingly require, a common construct set permits buyer and supplier to be assessed on the same dimensions and their results read against one another. Separate instruments by firm size make that comparison impossible.

6. Limitations and Planned Validation

The framework presented here is not empirically validated. That is the principal limitation of this paper.

Construction evidence is practitioner-derived and predominantly large-enterprise. Expert judgement rather than data currently supports the claim that the constructs are complete and that the stages discriminate. No claim is made that higher maturity produces better outcomes. The framework is offered as a diagnostic instrument, not as a tested theory of transformation success.

6.1 Planned validation

A two-stage validation study is in preparation.

- Expert stage: semi-structured interviews with ten to fourteen senior leaders holding direct accountability for multi-site smart manufacturing programs, including at least four from organizations below five hundred employees and at least two advisors working with small manufacturers. Experts will be asked to identify missing constructs before being shown the framework, then to assess construct completeness, stage discrimination and scale portability. Revisions and their rationale will be reported.

- Practitioner stage: a structured survey of thirty-six to forty-five respondents, stratified on two axes, technology domain and organization size band. The instrument will measure construct relevance, descriptor accuracy and advancement-criterion clarity, and will collect self-assessed current stage per dimension. A triangulation subset of six to eight organizations will provide a second respondent in a different role, with inter-rater agreement reported.
- Two self-reported outcome items will permit an exploratory association between maturity profile and outcome. The association will be correlational, single-informant and susceptible to common method variance [18]. It is not a test of predictive validity and will not be presented as one.

6.2 Limitations the planned study will not resolve

- Geographic and contextual boundary. Construction and planned validation are both situated in a North American context. Whether the constructs hold in European, Asian or other manufacturing contexts is not addressed here, and no claim is made about them. Cross-national validation is the first required extension.
- Assessor reliability. Agreement between respondents within an organization is not the same property as agreement between independent assessors applying the sixty criteria to the same organization. The second property is necessary if the framework is to be used by advisory networks. Criteria requiring demonstrated instances reduce assessor discretion, but that is a design mitigation rather than evidence. An assessor-reliability study is the second required extension.
- Validation evidence narrowed by measurement specification. Formative specification is appropriate to the artifact but removes access to internal-consistency and factor-analytic evidence, as set out in Section 3.2.
- Sample size at band level supports description rather than inference. No significance testing across size bands is planned at the intended sample size.

6.3 Further extensions

Longitudinal tracking across stage advancement is required to establish that maturity progression precedes outcome improvement. Extension to supply-network contexts, in which buyers and suppliers are assessed on comparable terms, follows from the design. Direct measurement of change absorption capacity, rather than assessment of it, would strengthen D3.

A further extension is domain scoping. The constructs are specified generally, but the failure modes they capture take concrete forms within particular technology classes. In artificial intelligence deployment, capability ownership becomes responsibility for model retraining and its continuity. Operational continuity becomes drift detection and monitored performance degradation against a baseline. Trust calibration becomes a central rather than a subsidiary concern. A domain-scoped variant would sharpen each construct into directly observable practice.

References

[1] Manufacturing Leadership Council, Smart Factories Survey, 2026.

[2] World Economic Forum and McKinsey & Company, Global Lighthouse Network reports on scaling digital manufacturing.

[3] McKinsey & Company, Industry 4.0: Capturing value at scale in discrete manufacturing, 2019.

[4] Vance, T., Jin, H., Price, C., Nimbalkar, S. and Wenning, T., "Smart manufacturing maturity models and their applicability: a review," Journal of Manufacturing Technology Management, Vol. 34, No. 5, 2023, pp. 735–770.

[5] Mittal, S., Romero, D. and Wuest, T., "Smart manufacturing and Industry 4.0 maturity models: implications for small and medium-sized enterprises," Journal of Manufacturing Systems, Vol. 49, 2018, pp. 194–214.

[6] Mittal, S., Romero, D. and Wuest, T., "Towards a smart manufacturing maturity model for SMEs," Advances in Production Management Systems, IFIP AICT, Springer, 2018.

[7] Schumacher, A., Erol, S. and Sihn, W., "A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises," Procedia CIRP, Vol. 52, 2016, pp. 161–166.

[8] Gökalp, E. and Martínez, V., "Digital transformation capability maturity model enabling the assessment of industrial manufacturing companies," Computers in Industry, Vol. 132, 2021.

[9] Hevner, A. R., March, S. T., Park, J. and Ram, S., "Design science in information systems research," MIS Quarterly, Vol. 28, No. 1, 2004, pp. 75–105.

[10] Peffers, K., Tuunanen, T., Rothenberger, M. A. and Chatterjee, S., "A design science research methodology for information systems research," Journal of Management Information Systems, Vol. 24, No. 3, 2007, pp. 45–77.

[11] Becker, J., Knackstedt, R. and Pöppelbuß, J., "Developing maturity models for IT management," Business and Information Systems Engineering, Vol. 1, No. 3, 2009, pp. 213–222.

[12] Kırmızı, M. and Kocaoğlu, B., "Digital transformation maturity model development framework based on design science: case studies in manufacturing industry," Journal of Manufacturing Technology Management, Vol. 33, No. 7, 2022.

[13] Jarvis, C. B., MacKenzie, S. B. and Podsakoff, P. M., "A critical review of construct indicators and measurement model misspecification in marketing and consumer research," Journal of Consumer Research, Vol. 30, No. 2, 2003, pp. 199–218.

[14] Diamantopoulos, A. and Winklhofer, H. M., "Index construction with formative indicators: an alternative to scale development," Journal of Marketing Research, Vol. 38, No. 2, 2001, pp. 269–277.

[15] Cohen, W. M. and Levinthal, D. A., "Absorptive capacity: a new perspective on learning and innovation," Administrative Science Quarterly, Vol. 35, No. 1, 1990, pp. 128–152.

[16] Bernerth, J. B., Walker, H. J. and Harris, S. G., "Change fatigue: development and initial validation of a new measure," Work and Stress, Vol. 25, No. 4, 2011, pp. 321–337.

[17] Lee, J. D. and See, K. A., "Trust in automation: designing for appropriate reliance," Human Factors, Vol. 46, No. 1, 2004, pp. 50–80.

[18] Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. and Podsakoff, N. P., "Common method biases in behavioral research," Journal of Applied Psychology, Vol. 88, No. 5, 2003, pp. 879–903.

[19] ISO 8000-1:2022, Data quality Part 1: Overview, International Organization for Standardization.

[20] ANSI/ISA-95 and IEC 62264, Enterprise-Control System Integration.

Data and Materials

The complete capability evidence criteria are published in Appendix A of this paper. No empirical data are reported in this preprint paper.

Declarations

The author declares no competing financial interests and received no funding for this work. The framework was developed independently, on the author's own initiative and outside the scope of any employment, and is not endorsed by, nor developed on behalf of, any employer, institute or standards organization. Practitioner observations informing its construction are reported in aggregate and anonymized form, and no proprietary, confidential or organizationally identifying information is disclosed. Prior practitioner publications by the author bearing on this work are characterized in Section 2.6.

A large language model was used for structuring and organization of the argument, for identifying candidate literature for review, and for copy-editing. All text in this paper was written by the author. The conceptual framework, the five capability constructs, the scale-banding design and the sixty evidence criteria originate in the author's professional practice and are the author's own work. All cited sources were independently located, read and verified by the author for existence and accurate characterization.

Appendix A : Capability Evidence Criteria

Twenty stage advancements specified across three organization size bands

A.1 Assessment rules

- All stated criteria for an advancement must hold. Partial satisfaction does not advance the stage: an organization meeting some proactive criteria but not all is assessed at collaborative.
- Assess against the size band of the operating unit holding technology decision authority, not corporate headcount. A three-hundred-person division with independent decision rights is assessed as mid-market within a twenty-thousand-person enterprise.
- Assess each dimension independently. Organizations commonly sit at different stages across dimensions, and the profile is more informative than any composite. Composite scoring is not recommended and forms no part of the framework.
- Criteria are stated as observable facts, not intentions. “A procedure exists” is observable; “the organization values documentation” is not. Where a criterion asks whether something was done, evidence of the instance is required.
- The capability assessed is identical across bands. Only the evidence demonstrating it differs. Applying large-enterprise evidence criteria to a small manufacturer measures headcount rather than capability.

A.2 D1: Capability Ownership & Continuity

Construct: whether the capability to operate and adapt deployed technology is held within the organization and persists independently of any individual or external party, and when accountability for post-deployment outcomes is established.

Advancement	Small (<100 employees)	Mid-market (100–499)	Large (500+, typically multi-site)
Reactive → Aware	A specific person is identified as responsible for the deployed system. The organization can name who to contact without referring to the vendor first.	Responsibility is assigned to a named role within an existing function. Day-to-day questions are not routed externally by default.	Internal resources are identified for each major system. External dependency is documented as a recognized risk.
Aware → Collaborative	The responsible individual has a protected time allocation. Written operating procedures exist and are sufficient for a successor. A backup person is identified by name.	Internal roles are defined with documented responsibilities. At least two individuals are trained on each critical system. Onboarding materials exist for the role.	Role definitions and a competency model are documented. At least one capability previously sourced externally is now held internally. Knowledge retention mechanisms exist independent of individuals.
Collaborative → Proactive	The backup has demonstrably operated the system, not merely been designated. Accountability for post-deployment outcomes was assigned to an operational owner before go-live. External parties are used for enhancement, not daily operation.	The internal team operates the system day-to-day without external dependency. Succession is documented for each critical role. Outcome accountability sits with the operating function and was settled before deployment.	Capability is institutionalized within a defined organizational unit. Competency development and succession are actively managed. The receiving site or function holds outcome accountability, established before go-live.
Proactive → Adaptive	The organization has transferred its operating approach to at least one external party — a supplier, peer, or customer — or trains others in it.	Internal capability generates improvements not specified by the vendor. The approach is documented in a form transferable to another unit or organization.	Capability is exported across the enterprise and externally. Internal innovation has produced methods or tooling adopted beyond the originating unit.

Table A.1. D1 capability evidence criteria by organization size band.

A.3 D2: Investment Governance & Sequencing

Construct: whether technology investment decisions are made against stated criteria, sequenced deliberately rather than concurrently, and standardized across the organization's operating units.

Advancement	Small (<100 employees)	Mid-market (100–499)	Large (500+, typically multi-site)
Reactive → Aware	A specific person decides technology investments. Decisions are not purely reactive to vendor approach or to whoever asks most persistently.	A recurring forum or named individual owns technology investment decisions across operating units.	Decision rights for technology investment are documented, even where not yet enforced consistently across sites.
Aware → Collaborative	A single documented decision point exists. Stated selection criteria are written down and applied. No more than one major initiative is in flight at a time.	An intake and prioritization process operates across lines or departments. Sequencing decisions are recorded. Standards are drafted, even if not yet enforced.	Portfolio intake is defined. Cross-site standards are documented. Prioritization is aligned across regions or business units.
Collaborative → Proactive	Selection criteria have been applied to at least one rejection, not only to approvals. A superseded tool has been retired on plan. No more than one major change is in flight per operating area.	Standards are enforced: at least one initiative has been modified or blocked for non-compliance. Concurrent initiative load is assessed against the operational calendar before scheduling.	Portfolio de-confliction operates across sites. Enterprise standards are enforced with a documented exception process. Sequencing accounts for assessed absorption load at each receiving site.
Proactive → Adaptive	Sequencing decisions are demonstrably informed by the outcomes of prior deployments, not only by resource availability.	Governance adapts: selection or sequencing criteria have been revised based on realized outcomes of earlier decisions.	Governance is self-correcting and the approach is transferable externally. Sequencing optimizes against measured absorption capacity rather than assumed capacity.

Table A.2. D2 capability evidence criteria by organization size band.

A.4 D3: Change Enablement & Workforce Readiness

Construct: whether the workforce is prepared, enabled and measured in its capacity to adopt and sustain a technology change, comprising readiness, change absorption capacity, and trust calibration.

Advancement	Small (<100 employees)	Mid-market (100–499)	Large (500+, typically multi-site)
Reactive → Aware	Training is delivered before go-live and goes beyond vendor handover. Someone is responsible for delivering it.	A structured training program exists and has a named owner. Training is not a single event at deployment.	A training or change function is formally engaged in deployments. A change management role is identified for each major initiative.
Aware → Collaborative	Someone verifies before go-live that users can perform the required tasks. Usage is checked after deployment. Other changes already underway are deliberately considered before scheduling.	Readiness is assessed pre-deployment against stated criteria. Adoption is tracked post-deployment. In-flight changes elsewhere are reviewed before a deployment date is set.	Readiness assessment gates deployment. Adoption is tracked at system level. Concurrent-change load is assessed across receiving sites before scheduling.
Collaborative → Proactive	Adoption is observed rather than assumed — actual usage is checked. At least one deployment has been deferred or resequenced because of change load. A superseded tool has actually been removed from use.	Behavioural adoption measures exist with defined thresholds that trigger intervention. There is a documented instance of resequencing due to absorption limits. Tool retirement is enforced, not encouraged.	Behavioural adoption signals feed defined intervention triggers. Change absorption capacity is a governed input to deployment scheduling. Retirement of superseded tools is enforced as policy.
Proactive → Adaptive	Where systems generate recommendations, the organization knows whether operators act on them, and has investigated why when they do not.	Trust calibration is addressed explicitly for recommendation-generating systems. Absorption capacity is actively replenished — deliberate quiet periods are planned, not merely observed.	Trust calibration is managed as a governed construct for autonomous and AI-driven systems. Absorption signals inform portfolio governance. Capacity replenishment is planned deliberately.

Table A.3. D3 capability evidence criteria by organization size band.

A.5 D4: Operational Continuity of Deployed Systems

Construct: whether availability, performance and recoverability of deployed digital assets are governed rather than left to chance or to the vendor.

Advancement	Small (<100 employees)	Mid-market (100–499)	Large (500+, typically multi-site)
Reactive → Aware	Someone is responsible when the system fails. A current vendor support contact exists and is known to the people who would need it.	A support model is defined with stated response expectations. Escalation has a named path.	Support model and escalation are defined across sites. Response expectations are documented and communicated.
Aware → Collaborative	A recovery procedure is documented. Backups are verified to exist, not assumed. Response expectations are agreed with the vendor in writing.	Continuity requirements are defined at design stage rather than after deployment. Recovery has been tested at least once. Monitoring is in place for critical systems.	Continuity is designed into deployment. Recovery is tested on a documented schedule. Monitoring covers systems identified as critical.
Collaborative → Proactive	Recovery is tested at a stated interval, not once. Availability problems are detected before users report them.	Monitoring is proactive with alerting. Continuity protocols are enforced rather than intended. Availability is tracked against a stated expectation.	Asset management is proactive and, where applicable, predictive. Continuity protocols are enforced. Uptime is governed with named accountability.
Proactive → Adaptive	Failure patterns are analysed and addressed at root cause rather than resolved case by case.	Common failure modes are remediated automatically or self-correct. Recurrence is measured and declining.	Failure is detected predictively. Continuity is a design default for new deployments rather than an addition to them.

Table A.4. D4 capability evidence criteria by organization size band.

A.6 D5: Value Realization & Sustainment

Construct: whether projected value is baselined, tracked after deployment rather than at it, and converted into subsequent investment decisions.

Advancement	Small (<100 employees)	Mid-market (100–499)	Large (500+, typically multi-site)
Reactive → Aware	A business reason for the investment is stated in writing before purchase.	A business case is required before technology investment proceeds.	A business case standard exists. Value categories are named and used consistently.
Aware → Collaborative	A baseline measure is captured before deployment. A single post-deployment review is scheduled at a defined interval.	Baseline and realization are tracked per initiative against a common value taxonomy. Variance is reviewed by the initiative owner.	A common value taxonomy and business-case template are in use. Realization checkpoints are defined. Portfolio-level tracking has been initiated.
Collaborative → Proactive	The post-deployment review was actually conducted and recorded for the most recent deployment, and its result influenced a subsequent decision.	Realization is measured at defined intervals after go-live rather than at it. At least one reinvestment or discontinuation decision is traceable to realization data.	Portfolio-level realization tracking operates. Realized-versus-projected variance is a governed metric. Reinvestment decisions are traceable to realization data.
Proactive → Adaptive	A decline in realized value has been noticed and acted on before the next budget cycle.	Realization is monitored closely enough to detect decay. Corrective action has been taken on at least one decaying initiative.	Realization visibility is continuous. Value decay is detected and arrested before it compounds. The value model is recalibrated from observed outcomes.

Table A.5. D5 capability evidence criteria by organization size band.

A.7 Design rationale for the banding

Three principles govern how criteria differ across bands. Stating them explicitly allows the banding to be assessed as principled rather than ad hoc, and allows practitioners to extend the criteria to cases not anticipated here.

Structure substitutes for documentation as scale increases

A small manufacturer demonstrates continuity through written procedure and a named backup, because no organizational structure exists to carry the capability. A large enterprise demonstrates the same capability through a defined unit with succession management, because documentation alone cannot survive the personnel churn of a large organization. Neither is a weaker form of the other.

Instance evidence substitutes for systemic evidence at small scale

Where a large enterprise demonstrates a governed process, a small manufacturer demonstrates a specific instance in which the behaviour occurred: one rejection against stated criteria, one deferral for change load, one review that changed a decision. A small organization performing the behaviour once deliberately exhibits the capability; requiring a documented process would measure bureaucracy rather than capability.

Formality rises with the number of people who must coordinate

Criteria become more formal across bands because coordination cost rises with headcount, not because larger organizations are held to a higher standard. A five-person decision requires no forum; a fifty-person decision does. The capability of deciding against stated criteria is constant.

— End of preprint —