BEYOND THE 70% MYTH: WHAT DO WE ACTUALLY KNOW ABOUT FAILURE RATES IN IMPROVEMENT AND TRANSFORMATION INITIATIVES?

DOI: 10.5937/JEMC2600001W

UDC:
Original Scientific Paper

André de WAAL[1]

[1]HPO Center, Oude Enghweg 2 1217 JC Hilversum The Netherlands
Corresponding author. E-mail: andredewaal@planet.nl
ORCID ID (https://orcid.org/0000-0002-6442-8642)

Paper received: 31.01.2026.; Paper accepted: 12.02.2026.

This paper examines what is actually known about failure rates in improvement and transformation initiatives, and why the headline '~70% failure claim' travels across domains. An integrative evidence synthesis maps failure-rate claims in academic and managerial literatures. Each claim is coded by failure type, unit of analysis, denominator, time horizon and rate form, and appraised for evidential robustness. Reported rates vary widely and are not comparable unless definitions, denominators and horizons are specified. Much academic repetition of '~70%' relies on secondary citation or underspecified claims rather than recalculated rates. Managerial benchmarks frequently define 'failure' as the complement of a stringent top-box success threshold and/or as non-sustainment over multi-year horizons, yielding higher apparent failure prevalence than narrower endpoints. Overall, the evidence does not support treating '70% failure' as a universal planning parameter; failure prevalence is best understood as a distribution conditional on what is counted and for how long. This study reframes the '70% debate' as a measurement and reporting problem, integrates academic and managerial evidence using a failure-type taxonomy and robustness appraisal, and proposes a 'Failure-Rate Reporting Minimum' to make future claims more falsifiable and practically useful.

Keywords: Failure rate; Organizational change; Transformation; Continuous improvement; Measurement.

ONLINE FIRST

INTRODUCTION

Many academic and managerial texts state that "around 70%" of change, improvement, or transformation initiatives fail (Heracleous & Bartunek, 2020; Jones et al., 2021; Moon et al., 2022; Vanhengel et al., 2026). High rates of reported failure in organizational change are frequently invoked in both academic and managerial literatures to motivate research agendas, justify adoption of structured approaches and governance mechanisms, and to frame theoretical inquiry into why change fails (Errida & Lotfi, 2021; Ferede et al., 2024; Monferdini & Bottani, 2024). Yet, this 70% often functions as a rhetorical anchor rather than a replicable statistic, and it is frequently unclear: what counts as an initiative (project vs program vs enterprise transformation), what counts as failure (no improvement, no sustained improvement, cost overrun, missed objectives, no statistically significant effect, termination), and what is the denominator and time horizon (6 months vs >3 years; pilot vs scale)? This matters because treating the 70% as a portable headline number invites false equivalence across fundamentally different phenomena. When the underlying unit of analysis, outcome criteria, denominator, and time horizon are left unspecified, the claim becomes non-falsifiable: it cannot be replicated, compared across studies, or meaningfully aggregated in reviews and meta-analyses. In effect, the same label ('failure') is applied to incompatible constructs, so subsequent theorizing risks explaining variation that is partly an artifact of inconsistent definitions and measurement rather than genuine differences in change dynamics. In addition, the rhetorical use of a single, decontextualized failure rate can shape managerial action in unhelpful ways, either normalizing poor

ISSN 2217-8147 (Online)
©2026 The Author(s). Published by University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

execution ("failure is expected") or triggering excessive control, compliance-heavy governance, and risk aversion that prioritizes short-term milestone delivery over learning, adaptation, and sustainment. In this light, it is no wonder that Cândido and Santos (2015) state that commonly cited failure figures are controversial and often based on evidence that is outdated, fragmentary, fragile or just absent, and that they therefore call for the true rate to be determined.

To heed this call, our research aims to answer the following research questions:

RQ1: What failure-rate claims are reported in the academic and managerial literatures on improvement and transformation initiatives?
RQ2: How do reported rates vary by failure type, unit of analysis, and time horizon?
RQ3: What is the evidential robustness of failure-rate claims (transparent denominator vs cited meme vs perception)?
RQ4: How do academic and managerial literatures differ in their implied 'failure-rate reality,' and what explains the divergence?

Answering the research questions is expected to yield several contributions. Theoretically, the study will clarify the failure-rate construct by showing that failure in change and improvement initiatives is not a single, binary outcome but a bundle of distinct constructs that are often conflated. Building on this clarification, it will propose a measurement framework: a minimum reporting standard for failure-rate claims that requires explicit specification of the failure definition, denominator, evaluation horizon, outcome type, and sustainment window. Finally, the study will develop a reconciliation model to explain why reported failure rates diverge across academic and managerial texts, for instance due to selection effects, differing definitions and incentives, and variation in evaluation designs. Practically, these contributions replace the '70% meme' with a diagnostic lens that asks, concretely, which type of failure is being prevented in a given initiative. This enables managers to apply a failure-mode risk assessment: designing different governance and control mechanisms for non-achievement (missed objectives), no-effect (no measurable improvement), and non-sustainment (improvements that do not persist). The overall managerial implication is an evidence-based stance: 70% should not be treated as a planning parameter, but as a context- and definition-dependent claim that must be interrogated before it is used to shape decisions.

The remainder of this article is structured as follows. Section 2 describes the research approach and appraisal method used to assess the evidential robustness of failure-rate claims. Sections 3 and 4 synthesize the academic and managerial literatures, respectively, and Section 5 compares these streams and discusses why headline rates diverge across definitions and evidence bases. Section 6 concludes by answering the research questions, noting limitations, and outlining directions for future research.

RESEARCH APPROACH

This study used an integrative evidence synthesis design to map, compare, and critically appraise reported failure-rate claims in two literatures that strongly shape the 70% narrative (Khalil et al., 2025): academic, peer-reviewed research (empirical studies and reviews that mention, cite, or calculate failure/success prevalence) and managerial/benchmark sources (consultancy reports, professional bodies, practitioner surveys, and industry benchmarks that report or popularize rates). Methodologically, this approach is best characterized as a scoping review with critical appraisal: the aim is not to estimate one pooled true failure rate, but to clarify constructs, document the distribution of reported rates, and assess the evidential basis on which those rates rest (Campbell et al., 2023; Peters et al., 2021).

Design, Scope and Search Strategy

Both types of improvement initiatives (e.g., Lean, Six Sigma, patient safety programs) and broader transformations (e.g., enterprise change programs, digital transformations, restructuring) were deliberately included because the literature frequently mixes these categories while treating their reported failure rates as interchangeable (Heracleous & Bartunek, 2020). Rather than eliminating this mixing through narrow exclusion, the study design treated it as an empirical phenomenon to be documented and explained (Campbell et al., 2023). The unit of analysis for each included source was the failure-rate claim (explicit or implied), linked to the source's stated or inferable definition, denominator, and time horizon.

The review included sources that met three conditions. First, the source contained any quantified statement about failure/success prevalence: an explicit percentage, an implied

ONLINE FIRST

percentage (e.g., success rate reported without the complement), a range/bound, or a distribution. Second, the source allowed identification of a unit of analysis (initiative/project/program/transformation; trial/intervention; organization-level transformation portfolio; or respondent-reported initiative experience). Third, the source provided enough information to code at least one of the following: failure definition, time horizon, or denominator/sampling frame (even if only partially, which is then reflected in the robustness rating). Sources that only asserted "most changes fail" without any quantification were excluded from the failure-rate database but might be referenced narratively to illustrate rhetorical patterns.

For each included source, a standardized set of variables was extracted and coded to enable like-for-like comparison:

- source type and domain (academic vs managerial; sector/industry; country/region if stated)
- unit of analysis (initiative/project/program/transformation; intervention/trial; portfolio; respondent perception)
- failure definition and failure type (e.g., non-achievement; no-effect vs comparator; non-sustainment; value shortfall; termination)
- time horizon and sustainment window (e.g., 6 months, 12 months, >3 years; explicit sustainment criterion if present)
- failure rate (as stated) and rate form (explicit, implied complement, bounded range, distribution)
- denominator and sampling frame (number of initiatives; number of respondents; number of trials/interventions; audited dataset vs self-report)
- evidence source (survey, systematic review, trial synthesis, case series, consultancy analytics, secondary citation)

To address RQ3, each failure-rate claim was assigned an evidence robustness label indicating how directly the rate is supported by traceable data. For academic sources, claims were classified as:

- A: rate calculated from the study's own data with a transparent denominator and observable rule for failure
- B: rate primarily cited from elsewhere (secondary citation), without recalculation from the study's own data
- C: perceived failure rate (respondent belief/assessment rather than observed outcomes)
- D: rate asserted with unclear provenance or insufficient traceability to a denominator and operational definition.

For managerial sources a parallel, managerially realistic scale was used:

- M1: explicit percentage (or valid complement) with clear denominator, sampling frame, and rule
- M2: percentage reported with partial transparency (e.g., secondary summary, limited method detail)
- M3: rate derived from conditional thresholds/bounds or specialized definitions that constrain comparability
- M4: qualitative quantification ("fewer than half…") that signals magnitude but not a reproducible rate
- M5: 'meme/received wisdom' statement with no traceable basis within the document.

Because scoping and other 'big picture' evidence syntheses often included heterogeneous evidence sources, including grey/practitioner materials with different outlet controls and reporting norms, a dual robustness-tagging scheme was applied that maintained cross-source comparability while using appraisal logic appropriate to each evidence genre (Manietta et al., 2022; Peters et al., 2021).

Synthesis Strategy

The synthesis proceeded in three layers, aligned with the research questions. First, a descriptive mapping of what rates are reported (RQ1) was provided, summarizing the frequency and range of claims by source type, domain, and unit of analysis. Second, rates were stratified by failure type, time horizon, and unit of analysis to show how reported prevalence changes when definitions and horizons change (RQ2). Third, the distribution of robustness labels (A-D; M1-M5) was analyzed, to distinguish observed rates from cited rates and perception-based rates (RQ3). Importantly, a single universal pooled failure rate was not computed because doing so would reproduce the very construct mixing this paper critiques. Where summary statistics were presented, they were calculated only within clearly comparable bins (e.g., "no-effect rates in controlled evaluations of CQI interventions" vs "objective-attainment shortfall rates in executive surveys"), and they were explicitly framed as bin-specific descriptors, not as a general planning parameter. Finally, the academic and managerial mappings were integrated to develop a reconciliation

ONLINE FIRST

explanation of why the two literatures implied different 'failure-rate realities' (RQ4), focusing on definitional choices, selection effects, incentives, and evaluation designs.

Failure-Rate Reporting Minimum (minimum reporting standard)

To make any stated 'failure rate' replicable and comparable it was treated as reportable only if the source specified (or could be unambiguously coded for) the following items: failure definition/failure type, unit of analysis, denominator and sampling frame, evaluation horizon, sustainment window (where applicable), outcome domain, comparator/ counterfactual (where applicable), and evidence basis (observed/calculated vs perceived vs cited). This requirement was directly motivated by the heterogeneity in how failure is operationalized across, for example, goal-attainment framings in the change literature (Heracleous & Bartunek, 2020), effectiveness framings in CQI trial syntheses (Hill et al., 2020), outcome-indicator framings in QIC reviews (Atkins et al., 2023), sustainment framings in hospital-wide improvement reviews (Moon et al., 2022) and care-home follow-up studies (Devi et al., 2023), and value/termination framings in process-improvement project failure research (Lameijer et al., 2022).

THE ACADEMIC LITERATURE REVIEW

This chapter maps the failure-rate claims reported in the academic literature and critically appraises how robust those claims are when coded against transparent criteria (definition, denominator, time horizon, and - where relevant - sustainment window). Because 'failure' is operationalized in multiple, non-equivalent ways, the chapter first clarifies five recurring failure types and then synthesizes the best-supported academic failure-rate evidence accordingly.

Defining 'Failure'

A central reason why the 'around 70% failure' claim persists is that failure is often treated as a single, intuitive outcome, while the literature operationalizes it in multiple, non-equivalent ways. In this study 'failure rate' is therefore defined as: *the proportion of initiatives in a specified denominator that meet an explicitly stated failure criterion, assessed over an explicitly stated evaluation horizon and, where relevant, a sustainment window*. This implies that 'failure rate' is not one number but a family of rates, indexed to what counts as failure and when it is assessed. To make this plurality explicit - and to prevent construct slippage between improvement projects, change programs, and transformations - five failure types that recur in the sources included in the literature review are distinguished:

- *Failure Type 1 (non-achievement of objectives) defines failure as goal shortfall*: the initiative does not meet stated targets (including common managerial variants such as not meeting goals within time and/or budget) (Obondi, 2020). This goal-deviation framing is the default meaning of 'change failure' in parts of the organizational change literature, where failure is described as not achieving the expected outcomes of a change program (Heracleous & Bartunek, 2020).
- *Failure Type 2 (no-effect vs comparator) defines failure in evaluation terms*: the initiative shows no statistically significant and/or no practically meaningful improvement compared with a counterfactual (standard practice, a control condition, or an appropriate baseline trend). This definition is common in trial-based and systematic review traditions, where 'no difference' findings are treated as evidence of limited or absent effectiveness, even when implementation activity occurred (Hill et al., 2020).
- *Failure Type 3 (non-sustainment) defines failure dynamically over time*: an initiative may achieve initial gains, but those gains do not persist beyond an agreed period, or the intervention is not continued at follow-up. Sustainability-focused studies frequently operationalize success as continued use of program components and continued achievement of desired outcomes, implying failure as discontinuation or erosion after implementation (Devi et al., 2023; Manderson et al., 2025; Moon et al., 2022).
- *Failure Type 4 (value shortfall) defines failure as material under-delivery of anticipated benefits* (financial, operational, service, safety, experience), even when activities were executed and some targets may have been partially met. This benefits-realization logic is central in process-improvement project failure research, which explicitly distinguishes under-delivery of anticipated value from other endpoints (Lameijer et al., 2022).
- *Failure Type 5 (termination due to under-delivery) is a stricter event-based endpoint*: an initiative is classified as failed when it is terminated because it has fallen short of goals or

ONLINE FIRST

is expected to under-deliver value. Importantly, termination is not automatically failure; projects can be ended for anticipated or sensible reasons (e.g., changed requirements) without implying under-delivery (Lameijer et al., 2022).

These types are not mutually exclusive in practice, an initiative can miss objectives (Type 1), show no measured effect (Type 2), and/or fail to sustain gains (Type 3). However, they are analytically distinct endpoints that require different denominators, horizons, and evidence standards. For that reason, the remainder of the review codes each reported 'failure rate' claim by failure type, unit of analysis, denominator/sampling frame, and time horizon (and sustainment window where applicable), using the Failure-Rate Reporting Minimum as the coding backbone.

The Best-Supported Academic Failure-Rate Evidence

This section focuses on the best-supported academic failure-rate evidence: quantified claims that can be classified by both failure type (what 'failure' means) and evidence robustness (how directly the rate is supported by traceable data). To make robustness comparable across academic sources, we apply a four-level tagging scheme: A (calculated from the study's own data; authors computed the rate from their dataset), B (cited estimate; authors reported a rate from prior literature/consultancy/other studies), C (perceived rate; survey respondents' perceptions/estimates), and D (rate asserted with unclear provenance; i.e., without evidential basis / unclear provenance). We interpret rates only within comparable failure-type logics rather than producing a single pooled 'average failure rate' that would reproduce construct mixing.

Table 1: Overview academic failure rates

Study	Failure rate reported	Label	Failure type(s)
Jones et al. (2021)	27.5% unsuccessful change programs (55 out of 200)	A	1
Hill et al. (2020)	54.2% no sig. difference on clinical process outcomes; 64.7% no sig. difference on patient outcomes; 100% no sig. difference on other outcomes	A	2
Burnes and Jackson (2011)	70%	B	1
Cândido and Santos (2015)	28%–90% (range of published strategy-implementation failure estimates)	B	1, 4
Damschroder et al. (2009)	Two-thirds (stated as an estimate from prior literature)	B	1
de Waal and Goedegebuure (2015)	60%–80%	B	1
Henrique et al. (2021)	66%–90% (fail to sustain)	B	3
Lameijer et al. (2022)	80% (reported as Bain survey finding: projects failing to deliver anticipated value)	B	4
Lennox et al. (2023)	55% (implied 'not continued delivery' from a cited review)	B	3
Marshall (2014)	70%+	B	1, 3
McLean and Antony (2014)	60% (Six Sigma initiatives fail to achieve desired results; cited)	B	1
Moon et al. (2022)	70% (fail to sustain; cited)	B	3
Scheirer and Dearing (2011)	40%–60% "do continue in some form" (i.e., implies ~40%–60% do not continue)	B	3
Wetzel and Dievernich (2014)	Up to 70%	B	1
Antony et al. (2019)	Perceived distribution (bands) e.g., 23% of respondents perceive >30% fail; 24% perceive 11–20% fail; 14% perceive <5% failed projects	C	1, 3, 4
Heracleous and Bartunek (2020)	60%–70% (reported as commonly claimed; evidential basis questioned)	D	1
Moore (2018)	70% (reported as commonly cited/industry belief; evidential basis criticized as unclear)	D	1

ONLINE FIRST

Table 1 shows that, although the academic literature frequently repeats "around 70%," most quantified failure-rate statements are ***not*** computed from the authors' own datasets. Of the 17 academic sources in Table 1, only two are tagged **A** (rates calculated from the study's own denominator). By contrast, the largest share of entries are **B** (cited estimates), including multiple headline rates in the 60-80% region that are imported from prior literature, consultancy surveys, or earlier reviews, rather than recalculated. Table 1 also contains one perceived-rate study **C**, that reports a distribution of *respondent beliefs* rather than observed failure prevalence, and two **D** entries that explicitly flag the commonly claimed 60-70% figure as lacking a traceable evidential basis (i.e., not perception but unclear provenance). Substantively, Table 1 also illustrates why the failure rate cannot be treated as a single construct: the failure types differ across studies and those types imply different denominators, horizons, and inference standards. Accordingly, the appropriate conclusion is not that the 'true rate' is 70% but that empirically grounded failure prevalence varies substantially by definition and evaluation design. This is consistent with the earlier study of Hughes (2011). Therefore, subsequent analysis should treat failure rates as *type-specific* rather than a portable headline statistic. Thus, the analysis indicates that the frequently cited ~70% failure rate is not well supported by consistent empirical evidence. Depending on how failure is defined (unsuccessful program vs no-effect in trials), empirically grounded 'failure' can range from roughly ~30% to ~65%+, so treating 70% as a single universal figure is not justified.

Why 70% Persists in Academic Citations

Despite the weak evidential basis for a single, universal failure rate, the "around 70%" claim persists in academic writing because failure is often treated as a single intuitive outcome even though the literature operationalizes it in multiple non-equivalent ways. This definitional plurality creates fertile ground for a headline number to travel across domains and study traditions without being falsifiable, particularly when the underlying unit of analysis, denominator, and time horizon are left unspecified (Frankenhaus et al., 2023; Knottnerus & Tugwell, 2020).

A first driver is citation cascades and secondary citation. Many academic texts deploy 70% as a contextual opener to signal the scale of the problem, but cite it indirectly via review articles, conceptual papers, or practitioner-linked sources rather than returning to an original dataset. Over time, repeated secondary citation creates an 'authority-by-repetition' dynamic: the claim becomes treated as background knowledge, and the cost of tracing the provenance (and the definitional assumptions embedded in the original estimate) is rarely justified within the space and scope constraints of a single article (Baethge & Jergas, 2025; Smith & Cumberledge, 2020). A second driver is construct slippage, where a single number is used interchangeably across constructs that should be kept analytically distinct (Ackerman & Lucas, 2025; Gonzalez et al., 2021). In practice, 'failure' may refer to non-achievement of objectives (Type 1), no-effect in controlled evaluation (Type 2), non-sustainment after initial gains (Type 3), or value shortfall despite activity completion (Type 4). When these distinct endpoints are blurred under one umbrella term ("change fails"), a number like 70% can appear to generalize because it is no longer anchored to one operational definition, one denominator, or one evaluation horizon. A third driver is rhetorical and incentive compatibility. A high, memorable failure-rate statistic provides a convenient narrative hook: it motivates urgency, helps justify research agendas, and supports the adoption of structured governance and intervention approaches. In that role, 70% functions less as a measurement claim than as a problem-framing device. Importantly, once the statistic is used primarily for framing, the academic incentive to interrogate its measurement basis weakens, especially when the focal contribution of the paper lies elsewhere (e.g., theory development, a case study, or a framework) (Araújo Sensever et al., 2022; Hyland & Jiang, 2023).

THE MANAGERIAL LITERATURE REVIEW

The academic literature is not the only or even the primary carrier of the 'around 70% failure' narrative. A substantial share of how failure rates are communicated in practice comes from managerial and benchmark sources (consultancies, professional bodies, surveys, and practitioner reports), and those sources often circulate widely and feed back into academic writing through citation cascades. Moore's (2018) review, for example, notes that professional/trade organizations repeatedly cite a 70% figure without providing a traceable empirical basis, and that such claims can become difficult to validate because references are vague, circular, or untraceable. Accordingly, this study deliberately

ONLINE FIRST

reviews managerial sources alongside peer-reviewed academic sources. This is consistent with a scoping-review-with-critical-appraisal logic: the aim is not to declare one pooled true failure rate, but to document what rates are reported, how they are defined, and how strong their evidential basis is, recognizing that different outlet types have different reporting norms and controls (Campbell et al., 2023; Manietta et al., 2022; Peters et al., 2021).

To preserve comparability while respecting different reporting conventions, managerial failure-rate claims are coded using the same failure-type logic introduced earlier (Types 1-5), but are assigned a managerial robustness label (M1-M5) indicating how reproducible the reported percentage is given the information disclosed (e.g., denominator/sampling frame, rule for success/ failure, and transparency of method).

The Best-Supported Managerial Failure-Rate Evidence

Table 2 summarizes the sources in the managerial literature dataset that stated (or allowed calculation of) a failure rate or success rate relevant to change, transformation, or improvement initiatives. As in the academic literature review, a rate was treated as meaningfully comparable only when the source provided (or could be reliably coded for) the minimum information needed to interpret it (failure definition/type, denominator/sampling frame, evaluation horizon, and where relevant sustainment window, and evidence basis). Each entry in Table 2 is therefore coded for failure rate reported (explicit or implied), robustness label (M1-M5), and failure type(s). Regarding the robustness label, M1 indicates an explicit percentage (or valid complement) with a clear denominator, sampling frame, and rule; M2 indicates a percentage with only partial transparency (e.g., limited method detail or a secondary summary); M3 indicates a rate derived from conditional thresholds/bounds or specialized definitions that constrain comparability; M4 indicates qualitative quantification (e.g., "fewer than half…") that signals magnitude without a reproducible rate; and M5 indicates a meme/received-wisdom statement with no traceable basis within the document.

ONLINE FIRST

Table 2: Overview managerial failure rates

Study	Failure rate reported	Label	Failure type(s)
Allas et al. (2018)	80% fail to meet objectives.	M1	1
Jacquemont et al. (2015)	74% (implied) from 26% very/completely successful.	M1	1, 3
Forbes Insight (2014)	20% (one in five) executives felt attempts had failed.	M2	1
White et al. (2022)	~72% (implied) from 28% success-likelihood at the low end (28%→73% range).	M3	1
Basford et al. (2015)	26% (implied) from 74% success under the condition that key elements are all present.	M3	1, 3
Bucy et al. (2021)	>67% (implied) because < 1/3 report success on both performance and sustainment.	M3	1, 3
Armbruster et al. (2023)	88% (implied) because only 12% report sustaining goals >3 years.	M3	3
Johnston et al. (2017)	No exact % stated; fewer than half say most/all change efforts met goals and sustained over time (directionally implies a majority do not meet that bar).	M4	1,3
Anand and Barsoux (2017)	About three-quarters (≈75%) of change efforts flop.	M5	1

Table 2 shows that managerial failure-rate claims span a very wide range - roughly from ~20% at the low end to ~88% at the high end - largely because managerial sources operationalize 'failure' using different thresholds, time horizons, and (often) multi-criterion definitions of success. For instance, Forbes Insight (2014) reported that 20% of executives felt prior transformation attempts had failed (a perception-based or self-assessed failure claim). At the other extreme, Armbruster et al. (2023)'s report on sustaining transformation impact found that while a majority of respondents report achieving most or all goals, only 12% report sustaining those goals for more than three years, implying an 88% non-sustainment rate under that strict sustainment criterion (Failure Type 3). A cluster of widely cited managerial rates fall in the 60-80% failure band, but Table 2 makes clear that these are typically shortfall rates rather than 'complete collapses'. Table 2 also shows that some

managerial failure rates come from adjacent benchmarking domains - especially project management and IT project delivery - where the underlying construct differs materially from organizational transformation success. Finally, Table 2 highlights that some influential managerial sources provide approximate rather than exact rates (e.g., 'less than one-third'), which signals magnitude but is not directly reproducible as a single percentage without access to the underlying distribution (M4-type evidence in our robustness scale).

Taken together, Table 2 supports two overarching points. First, managerial failure-rate claims are highly sensitive to the success threshold (e.g., 'fully succeed' or 'very/completely successful' vs. 'met most goals'), and therefore can generate very different implied failure percentages even when describing broadly similar phenomena. Second, the table demonstrates why reporting the denominator, definition, and time horizon - the minimum needed for interpretability - is essential: otherwise, numbers in the same '70% neighborhood' can refer to fundamentally different endpoints (objective shortfall, benefits shortfall, lack of sustainment, or outright termination).

What Managerial Sources are Actually Measuring

A critical insight from Table 2 is that most managerial failure rates are not measuring failure as a binary, externally audited endpoint. Instead, they most commonly measure one (or a combination) of the following:

- *Top-box success shortfall (strict success thresholds)*. Many consultancy surveys define success using stringent criteria (e.g., 'very/completely successful', sometimes on multiple dimensions simultaneously). Under this approach, the failure rate is often the complement of a 'top-box success rate' (Mittal et al., 2023). One survey illustrates this clearly: it reports the share of respondents whose transformations are 'very' or 'completely' successful on both performance and organizational health, which implies a failure rate for not meeting that combined threshold even though many of those 'non-top-box' cases may still have delivered partial value.
- *Sustainment as a dominant driver of failure*. Some managerial sources explicitly incorporate sustainability or persistence over time, shifting the endpoint from immediate implementation success to durability (Failure Type 3). In one such study, the implied failure rate is largely driven by the difficulty of maintaining gains over multi-year horizons rather than by initial execution collapse: 56% report achieving most/all goals, but only 12% report sustaining for more than three years.
- *Objective attainment vs. benefits realization vs. full success*. Managerial sources frequently blend (or ambiguously separate) goal achievement, benefits realization, and holistic full success. One of those studies distinguishes 'met objectives', 'partially met', and 'did not meet', but secondary retellings often collapse these into a binary success/failure split by treating partial as failure, producing a higher implied failure rate than the original three-category framing necessarily warrants.
- *Benchmarking in adjacent domains with different denominators and endpoints*. Project management benchmarks and IT project delivery benchmarks often use different units of analysis (projects), different denominators (portfolios of projects, often in IT), and different outcome criteria (e.g., meeting business goals, challenged vs canceled). These are valuable as managerial performance signals, but they should not be treated as interchangeable with transformation failure rates unless the construct mapping is made explicit (Types 1/4/5 depending on interpretation).

In other words, the managerial literature tends to produce failure rates as performance shortfall indicators – often perception-based and threshold-dependent – rather than as controlled effectiveness estimates (Failure Type 2) or independently verified population prevalence estimates. This is not a flaw; it reflects the purpose and genre of managerial reporting. But it does mean that managerial rates should be used carefully in academic argumentation: they are best treated as evidence of how organizations perceive and narrate transformation outcomes, and as descriptive indicators of perceived underperformance under specific success definitions, rather than as directly poolable estimates of a single underlying true failure probability.

ONLINE FIRST

COMPARISON AND DISCUSSION

The previous Chapters show that the academic and managerial literatures do not merely report different estimates of failure rate, they report different phenomena under the same label. This chapter therefore compares the two evidence streams explicitly and discusses what the divergence means for interpreting the 70% failure claim, theory development on change and improvement outcomes, and managerial use of failure-rate statistics.

Why Academic and Managerial Rates Diverge

First, the unit of analysis differs. Academic sources frequently treat the unit as an evaluated intervention or as a coded case in a synthesis, whereas managerial sources often treat the unit as a transformation episode as perceived and categorized by respondents (or occasionally as a portfolio of initiatives). This matters because 'initiative', 'change program', and 'transformation' are not interchangeable denominators: they differ in scope, duration, governance complexity, and what is considered a reasonable success threshold. Second, the endpoint (i.e., what counts as failure) differs. Academic traditions often operationalize failure as no-effect versus a comparator (Failure Type 2) or as non-sustainment (Failure Type 3) when follow-up is available; managerial traditions more often operationalize failure as objective shortfall (Type 1) or value shortfall (Type 4), and frequently do so using stringent success categories (e.g., 'very/completely successful' or 'success on multiple dimensions'). When the endpoint shifts from 'no statistically/practically meaningful effect' to 'not top-box success', the implied failure rate will typically rise even when many initiatives deliver partial value.

Third, the evaluation standard differs (i.e., how success/failure is established). In the academic literature, the most robust rates (label A) are computed from an explicit denominator and a defined outcome rule. Much of what appears as 70% in academic writing, however, is imported through secondary citation (label B) or asserted without traceable provenance (label D). In managerial sources, many rates are derived from surveys and benchmarks in which success is self-reported and threshold-dependent (e.g., multi-criterion success definitions, conditional success under specific enabling conditions), yielding robustness levels M2-M3 more often than fully reproducible M1 rates. The result is that managerial rates commonly describe perceived performance shortfall under a strict success definition, while academic A-type rates describe observed outcomes under an evaluation definition. Fourth, incentives and rhetorical function differ. Managerial benchmarks are designed to motivate action, support prioritization, and legitimate investment in structured approaches; they therefore have a natural incentive toward a memorable, urgency-inducing framing. Academic writing also uses problem-framing devices, but the incentive structure differs: a paper's contribution often lies elsewhere (theory, mechanism, method, or a framework), so authors may adopt a familiar headline statistic as context without spending scarce article space to fully trace provenance and definitional assumptions (Trueblood et al., 2025). This 'framing use' is consistent with why citation cascades persist, and why the same number can travel even when it is not treated as a testable measurement claim (Frankenhuis et al., 2023; Pavlovic et al., 2022).

In conclusion, the divergence between the two types of literature is not a conflict to be resolved by choosing which number is right. Rather, it reflects a systematic mismatch in units, endpoints, horizons, and evidence bases. The same percentage neighborhood (e.g., around 70%) can therefore refer to fundamentally different underlying constructs.

Key Empirical Takeaway: 70% is Not a Universal Parameter

Across both literatures, the evidence does not support treating '70% failure' as a portable planning parameter for change in general. The defensible conclusion from the mapped evidence is that failure prevalence is a distribution, not a constant, and that the distribution shifts predictably with: failure type (objective shortfall vs no-effect vs non-sustainment vs value shortfall vs termination); time horizon and sustainment window (immediate outcomes vs multi-year persistence); evaluation design (controlled comparison vs uncontrolled before-after vs perception-based surveys); unit and scale (single projects vs programs vs enterprise transformations; portfolio vs single initiative); and success threshold (top-box or multi-criterion 'full success' vs 'met most goals' vs 'some gains'). In other words: when 'failure' is defined narrowly (e.g., termination due to under-delivery) the rate will usually be lower than when 'failure' is defined as 'not fully successful on multiple dimensions for >3 years'. Both may be

ONLINE FIRST

legitimate but they should not be treated as the same construct.

Theoretical Consequences

There are several consequences for the theory. First, binary success/failure has to be replaced with failure-mode theory. The literature mapped here implies that a single binary outcome ('success vs failure') collapses distinct phenomena that likely have different causal mechanisms and different remedies. A no-effect outcome (Type 2) has different theoretical antecedents (problem selection, theory of change, fidelity, measurement sensitivity, counterfactual validity) than non-sustainment (Type 3), which implicates routinization, ownership, turnover, governance continuity, and feedback systems. Likewise, value shortfall (Type 4) invites benefits-realization and capability explanations that are not captured by effectiveness framings. Second, shift from static endpoints to trajectories. Both literatures implicitly point to trajectory patterns - improve → stall → adapt → sustain (or erode) - but headline rates obscure those dynamics. A trajectory perspective (with explicit time horizons) would allow theory to distinguish early implementation success from long-term sustainment, and to explain when and why improvements decay even after initial gains. Third, reframe failure-rate claims as measurement claims. A central theoretical correction is to treat any failure rate as a measurement statement that requires specification (definition, denominator, horizon) rather than as generic background knowledge. Without that, subsequent theorizing risks explaining variance that is partly an artifact of inconsistent operationalization rather than genuine differences in change dynamics.

Practical Consequences

The main consequence for practice is the managerial risk, not that leaders see a high failure rate and become motivated, but that they use a decontextualized rate to adopt the wrong control logic. Thus, the practical implication is to move from fatalism ("most fail") to failure-mode risk management, where governance choices are matched to the dominant failure risk:

- If the dominant risk is no-effect (Type 2): emphasize problem selection, a plausible theory of change, piloting, measurement validity, and counterfactual thinking. Invest early in defining what 'effect' means (process/outcome/financial), what baseline trend is expected, and what magnitude is meaningful (consistent with continuous-improvement evidence that measurement mechanisms and specific objectives are prerequisites for sustain improvement, see Spasojević Brkić et al. (2020).
- If the dominant risk is non-sustainment (Type 3): design for embedding and durability, i.e., ownership, routinization, turnover-proofing, ongoing measurement and feedback loops, and clarity on the sustainment window (what must persist, for how long, and with what tolerance for drift).
- If the dominant risk is objective shortfall/value leakage (Types 1 and 4): strengthen prioritization, resourcing, role clarity and middle-management change agency (Kreiner et al., 2018), benefits realization practices, and follow-through governance. In practice, many failures in managerial surveys may be not top-box success, meaning partial value exists; the managerial task is then to recover value and learn, rather than to treat the initiative as a total loss.

A key managerial takeaway is therefore: do not ask "How do we beat 70% failure?" but ask: "Which failure type are we trying to prevent, in which unit, over what horizon, with what evidence standard?" The answer determines the appropriate governance and intervention approach.

A Proposed Reporting Checklist: the 'Failure-Rate Reporting Minimum'

To prevent construct mixing and to make failure-rate claims replicable and comparable, the proposal is that any academic or managerial source stating a failure (or success) rate should report at minimum the following items (or be treated as non-reproducible if it does not):

- failure definition / failure type (Types 1–5) and whether partial success is treated as failure
- unit of analysis (project, initiative, program, transformation, intervention/trial, portfolio)
- denominator and sampling frame (N, sector, geography, inclusion criteria; audited vs convenience sample)
- evaluation horizon and, where relevant, sustainment window (e.g., 12 months; >3 years)
- outcome domain (process, outcome/impact, financial, capability/health; single vs multi-criterion success)
- comparator/counterfactual (where applicable) and what counts as 'no-effect'

ONLINE FIRST

– evidence basis: observed/calculated from own data vs perceived/self-report vs modelled vs secondary citation
– rate form: explicit rate vs implied complement vs bounded/thresholded claims (and how the complement was computed)

Applying this minimum does not police managerial benchmarks; it simply makes clear what a rate means, what it can be compared with, and what it cannot. It also creates a clean bridge between academic and managerial reporting: both can coexist, but only if their claims are explicitly indexed to definitions, denominators, and horizons.

CONCLUSION, LIMITATIONS, AND FUTURE RESEARCH

This Chapter closes the article by synthesizing what the evidence map implies for interpreting widely cited failure-rate claims and by answering the research questions. It then acknowledges the study's limitations and sets out a focused agenda for future research to make failure-rate reporting more comparable, falsifiable, and practically useful.

Conclusion

The research questions can now be answered.

RQ1: What failure-rate claims are reported in the academic and managerial literatures?
The mapping conducted in this article shows that both literatures frequently circulate failure-rate claims in the 'around 70% neighborhood', but the empirical situation is more heterogeneous than the headline suggests. In the academic dataset (Table 1), quantified claims span from clearly calculated rates such as 27.5% unsuccessful change programs (Type 1) to substantially higher no-effect rates within controlled evaluation traditions (Type 2), while a large share of academic mentions of 60-80% appear as cited estimates rather than rates recalculated from the study's own denominator. In the managerial dataset (Table 2), reported or implied failure rates span a wide range (roughly ~20% to ~88%), with a visible cluster of widely cited claims between 60-80% that often reflect performance shortfall under strict success criteria rather than 'complete collapse'.

RQ2: How do reported rates vary by failure type, unit of analysis, and time horizon?
Variation is systematic and predictable once failure is treated as a family of constructs (Types 1-5) rather than one binary outcome. Rates are higher when the unit of analysis is a broad transformation as reported by executives rather than a clearly bounded intervention/trial, the success threshold is stringent (e.g., very/completely successful, often on multiple dimensions), and the time horizon incorporates a long sustainment window (e.g., >3 years). In managerial sources, this is visible in 'top-box shortfall' logic (failure = complement of a top-box success rate) and in sustainment-based definitions where long-horizon persistence drives the implied failure rate upward. In academic sources, no-effect vs comparator (Type 2) is not commensurate with non-achievement of objectives (Type 1) or non-sustainment (Type 3): each implies different denominators, inference standards, and interpretive meanings.

RQ3: What is the evidential robustness of failure-rate claims?
The robustness appraisal shows why a universal 'true rate' cannot be responsibly inferred from the literature as it is commonly cited. In the academic set (Table 1), only a small minority of claims are calculated directly from an author's own denominator (A), while most are secondary-citation claims (B), alongside perceived-rate evidence (C) and uncertain-provenance claims (D). In the managerial set (Table 2), the evidence base is mixed: some sources provide reproducible percentages (M1), but many are partially transparent, threshold-bound, or only directionally quantified (M2-M4), and a subset functions as received wisdom (M5). The implication is not that managerial sources are wrong but that they often measure a different endpoint (e.g., perceived shortfall against strict criteria), and disclose different levels of methodological detail.

RQ4: What explains the divergence in the academic and managerial 'failure-rate realities'?
The divergence is explained by four structural differences: units (bounded interventions/trials vs broad transformations/portfolios), endpoints (no-effect vs comparator vs objective shortfall vs sustainment vs value realization), evaluation standards (peer-reviewed designs vs perception-based surveys/benchmarks), and incentives and genre purposes (scholarly contribution vs managerial persuasion/benchmarking). Together these differences create the conditions under which a single headline number can travel across domains even when it is not used as a testable measurement

ONLINE FIRST

claim: the statistic becomes a framing device, while the definitional and denominational assumptions that would make it falsifiable remain unspecified.

The overall conclusion is that the literature does not support treating '70% failure' as a universal planning parameter. The defensible conclusion is that failure prevalence is a distribution that depends on failure type, unit of analysis, time horizon, and evidence basis, and that future work should report failure-rate-by-type using transparent denominators and explicit horizons (the Failure-Rate Reporting Minimum).

Limitations

This study is an integrative evidence synthesis with critical appraisal, not an attempt to estimate a pooled 'true' failure rate; that design choice is deliberate, because pooling across non-equivalent definitions would reproduce construct mixing. This brings with it several limitations. First, coverage is necessarily incomplete, especially for managerial and grey literature where access, searchability, and methodological disclosure vary widely. The dataset is therefore best interpreted as a mapped sample of influential and codable failure-rate claims rather than an exhaustive census. Second, many managerial rates rely on self-reporting, which can introduce recall effects, social desirability bias, and respondent interpretation variance; these sources are valuable for understanding perceived shortfall under specific thresholds, but they are not equivalent to audited prevalence. Third, cross-domain comparison risks construct drift: 'transformation', 'change program', 'QI initiative', and 'process improvement project' can differ materially in scope, governance, and feasible horizons; our coding framework mitigates this by distinguishing failure types and units, but residual ambiguity remains in some sources. Finally, robustness labels (A-D; M1-M5) and failure-type coding involve judgement when sources are underspecified; while this judgement is structured and rule-based, different reviewers might code a subset of borderline cases differently.

Future Research Agenda

A useful next phase of research is to move from 'headline debate' to measurement infrastructure and trajectory evidence, by undertaking the following research avenues. First, build a cross-sector, open failure-rate evidence base using shared definitions. For this, create a repository (or at least shared templates) where initiatives are recorded with the Failure-Rate Reporting Minimum. This would enable comparability without forcing artificial pooling. Second, retrospective surveys and post-hoc estimates should be complemented with prospective tracking of improvement and transformation portfolios, measuring not only initial outcomes but persistence (e.g., 12, 24, 36+ months). This is essential to separate early implementation success from non-sustainment (Type 3). Third, create trajectory-based outcome models rather than binary endpoints. For this, study outcome trajectories (improve → stall → adapt → sustain/erode) and link them to governance, capability, and context variables. This would yield more actionable theory than treating all 'non-top-box' outcomes as homogeneous failure. Fourth, develop and validate a failure-mode instrument. Design a measurement instrument that predicts distinct failure modes (Type 1 vs Type 2 vs Type 3 vs Type 4/5), so that research and practice can target mechanisms that plausibly drive each mode (e.g., problem selection and measurement validity for Type 2; embedding and turnover-proofing for Type 3). Fifth, map the citation network behind the 70% claim to show how it mutates across outlets, which sources act as hubs, and how definitional assumptions drop out during repetition. This would turn the 'meme diagnosis' into testable, cumulative evidence.

Taken together, these directions replace the question "Is the true failure rate 70%?" with a more useful research program: "Which failure mode, for which unit, over which horizon, under which evaluation standard and with what predictable drivers?"

REFERENCES

Ackerman, L.S., & Lucas, R.E. (2025). What's in a name? Exploring overlap among self-belief constructs. *Journal of Research in Personality*, *117*, 104631. https://doi.org/10.1016/j.jrp.2025.104631

Allas, T., Checinski, M., Dillon, R. Dobbs, R., Hieronimus, S., & Singh, N. (2018), *Delivering for citizens: How to triple the success rate of government transformations*, McKinsey & Company.

Anand, N., & Barsoux, J-L. (2017). What Everyone Gets Wrong About Change Management. *Harvard Business Review*, November-December, 80-85

Antony, J., Lizarelli, F.L., Fernandes, M.M., Dempsey, M., Brennan, A., & McFarlane, J. (2019). A study into the reasons for process improvement project failures: results from a pilot survey. *International Journal of Quality & Reliability Management*,

ONLINE FIRST

36(10), 1699-1720. https://doi.org/10.1108/IJQRM-03-2019-0093

Araújo Sensever, F. de; Lucena Alves, P. de, da S. Lima, G., Loomans, B., Opdam, N., & Pereira-Cenci, T. (2022). Spin and reporting in systematic reviews with meta-analysis of randomized clinical trials in restorative dentistry. *Journal of Dentistry*, *125*, 104282. https://doi.org/10.1016/j.jdent.2022.104282

Armbruster, S., Busellato, N., Robson, N., & Sakamoto, T. (2023). *How to implement transformations for long-term impact*, McKinsey & Company.

Atkins, E., Birmpili, P., Glidewell L, et al. (2023). Effectiveness of quality improvement collaboratives in UK surgical settings and barriers and facilitators influencing their implementation: a systematic review and evidence synthesis. *BMJ Open Quality*, *12*, e002241. https://doi.org/10.1136/bmjoq-2022-002241

Baethge, C., & Jergas, H. (2025). Systematic review and meta-analysis of quotation inaccuracy in medicine. *Research Integrity and Peer Review*, *10*(1), 13. https://doi.org/10.1186/s41073-025-00173-z

Basford, T., Schaninger, B., & Viruleg, E. (2015), *The science of organizational transformations*, McKinsey & Company

Bucy, M., Schaninger, B., VanAkin, K., & Weddle, B. (2021), *Losing from day one: Why even successful transformations fall short*, McKinsey & Company.

Burnes, B., & Jackson, P. (2011). Success and Failure In Organizational Change: An Exploration of the Role of Values. *Journal of Change Management*, *11*(2), 133-162. https://doi.org/10.1080/14697017.2010.524655

Campbell, F., Tricco, A.C., Munn, Z., Pollock, D., Saran, A., Sutton, A., White, H., & Khalil, H. (2023). Mapping reviews, scoping reviews, and evidence and gap maps (EGMs): the same but different- the "Big Picture" review family. *Systematic Reviews*, 12, article 45. https://doi.org/10.1186/s13643-023-02178-5

Cândido, C.J.F., & S.P. Santos (2015). Strategy implementation: What is the failure rate?. *Journal of Management & Organization*, *21*(2), 237-262. https://doi.org/10.1017/jmo.2014.77

Damschroder, L.J., Aron, D.C., Keith, R.E., Kirsh, S.R., Alexander, J.A., & Lowery, J.C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*, *4,* 50. https://doi.org/10.1186/1748-5908-4-50

Devi, R., Martin, G.P., Banerjee, J., Gladman, J.R., Dening, T., Barat, A., & Gordon, A.L. (2023). Sustaining interventions in care homes initiated by quality improvement projects: a qualitative study. *BMJ Quality & Safety*, *32*(11), 665-675. https://doi.org/10.1136/bmjqs-2021-014345

de Waal, A., & Goedegebuure, R. (2015). Verbeter het process van Procesverbetering" [Improve the process of process improvement], *Holland Management Review*, *160*, 25-31.

Errida, A., & Lotfi, B. (2021). The determinants of organizational change management success: Literature review and case study. *International Journal of Engineering Business Management*, *13*. https://doi.org/10.1177/18479790211016273

Ferede, W.L., Endawoke, Y., & Tessema, G. (2024). Effects of strategic leadership on change management: examining the mediating roles of accountability, knowledge management, and organizational culture in public organizations: a study in Central Gondar, Ethiopia, *Cogent Business & Management*, *11*(1, 2416613. https://doi.org/10.1080/23311975.2024.241661

Forbes Insight (2014), *Making The Change, Planning, Executing And Measuring Successful Business Transformation*, Forbes Insight & Oracle.

Frankenhuis, W. E., Panchanathan, K., & Smaldino, P. E. (2023). Strategic ambiguity in the social sciences. *Social Psychological Bulletin*, 18, Article e9923. https://doi.org/10.32872/spb.9923

Gonzalez, O., MacKinnon, D.P., & Muniz, F.B. (2021). Extrinsic Convergent Validity Evidence to Prevent Jingle and Jangle Fallacies. Multivariate Behavioral Research, *56*(1), 3-19. https://doi.org/10.1080/00273171.2019.1707061

Henrique, D.B., Filho, M.G., Marodin, G., Jabbour, A.B.L. de S., & Chiappetta Jabbour, C.J. (2021). A framework to assess sustaining continuous improvement in lean healthcare. *International Journal of Production Research*, *59*(10), 2885-2904. https://doi.org/10.1080/00207543.2020.1743892

Heracleous, L., & Bartunek, J. (2020). Organization change failure, deep structures and temporality: Appreciating Wonderland. *Human Relations*, *74*(2), 208-233. https://doi.org/10.1177/0018726720905361

Hill, J.E., Stephani, AM., Sapple, P., & Clegg, A.J. (2020). The effectiveness of continuous quality improvement for developing professional practice and improving health care outcomes: a systematic review. *Implementation Science*, *15*, 23. https://doi.org/10.1186/s13012-020-0975-2

Hughes, M. (2011). Do 70 Per Cent of All Organizational Change Initiatives Really Fail?. *Journal of Change Management*, *11*(4), 451-464. https://doi.org/10.1080/14697017.2011.630506

Hyland, K., & Jiang, F. (2024). Hyping the REF: promotional elements in impact submissions. *Higher Education*, *87*, 685-702. https://doi.org/10.1007/s10734-023-01030-y

Jacquemont, D., Maor, D., & Reich, A. (2015), *How to beat the transformation odds*, McKinsey & Company

Jones, J., Firth, J., Hannibal, C., & Ogunseyin, M. (2021). Factors Contributing to Organizational Change Success or Failure: A Qualitative Meta-Analysis of 200 Reflective Case Studies". In: I. Management Association (Ed.), *Research Anthology on Digital Transformation, Organizational Change,*

IN THE PRESS

and the Impact of Remote Work, IGI Global Scientific Publishing, 1427-1450. https://doi.org/10.4018/978-1-7998-7297-9.ch071

Johnston, A., Lefort, F., & Tesvic, J. (2017), *Secrets of successful change implementation*, McKinsey & Company.

Khalil, H., Welch, V., Grainger, M., & Campbell, F. (2025). Methodology for mapping reviews, evidence maps, and gap maps. *Research Synthesis Methods*, *16*(5), 786-96. https://doi.org/10.1017/rsm.2025.25

Knottnerus, J.A., & Tugwell, P. (2020). It's always about numerators and denominators (N/D). Journal of Clinical Epidemiology, *126*, A7-A9. https://doi.org/10.1016/j.jclinepi.2020.09.013

Kreiner, J., Stanković, M., Sajfert, D., Terek, E., & Škrinjarić, Z. (2018). The impact of middle managers in change management in the Serbian metal processing industry. *Journal of Engineering Management and Competitiveness*, *8*(1), 20-27. https://doi.org/10.5937/jemc1801020K

Lameijer, B.A., Antony, J., Borgman, H.P., & Linderman, K. (2022). Process improvement project failure: a systematic literature review and future research agenda. *International Journal of Lean Six Sigma*, *13*(1), 8–32. https://doi.org/10.1108/IJLSS-02-2020-0022

Lennox, L., Antonacci, G., Harris, M., & Reed, J. (2023). Unpacking the 'process of sustaining'-identifying threats to sustainability and the strategies used to address them: a longitudinal multiple case study. *Implementation Science Communications*, *4*(1, 68. https://doi.org/10.1186/s43058-023-00445-z

Manderson, K., Taylor, N.F., Lewis, A., Harding, K.E. (2025). Service-level interventions to reduce waiting time in outpatient and community health settings may be sustained: a systematic review. *BMJ Open Quality*, *14*, e003235. https://doi.org/10.1136/bmjoq-2024-003235

Manietta, C., Rommerskirch-Manietta, M., Purwins, D., & Roes, M. (2022). Consulting concepts and structures for people with dementia in Germany: a protocol for a 'grey-shaded' scoping review. *BMJ Open*, *12*(4, e059771. https://doi.org/10.1136/bmjopen-2021-059771

Marshall, D.A. (2014), *Lean Transformation: Overcoming the Challenges, Managing Performance, and Sustaining Success*, PhD Thesis, University of Kentucky.

McLean, R., & Antony, J. (2014). Why continuous improvement initiatives fail in manufacturing environments? A systematic review of the evidence. *International Journal of Productivity and performance Management*, *63*(3), 370-376. https://doi.org/10.1108/IJPPM-07-2013-0124

Mittal, V., Han, K., Frennea, C., Blut, M., Shaik, M., Bosukonda, N., & Sridhar, S. (2023). Customer satisfaction, loyalty behaviors, and firm financial performance: what 40 years of research tells us. *Marketing Letters*, *34*, 171-187. https://doi.org/10.1007/s11002-023-09671-w

Monferdini, L., & Bottani, E. (2024). How do businesses utilize change management for process optimization? A cross-analysis among industrial sectors. *Business Process Management Journal*, *30*(8), 371-414. https://doi.org/10.1108/BPMJ-03-2024-0158

Moon, S.E.J., Hogden, A., & Eljiz, K. (2022). Sustaining improvement of hospital-wide initiative for patient safety and quality: a systematic scoping review. *BMJ Open Quality*, *11*, e002057. https://doi.org/10.1136/bmjoq-2022-002057

Moore, M. (2018), *An Exploration of the Causes of Success and Failure of Managed Change*, Walden Dissertations and Doctoral Studies, no. 5229, Walden University, Minneapolis. https://scholarworks.waldenu.edu/dissertations/5229

Obondi, K. C. (2020). The impact of project risk audits on construction project success. *Journal of Engineering Management and Competitiveness*, *10*(2), 103–115. https://doi.org/10.5937/jemc2002103O

Pavlovic, V., Weissgerber, T., Stanisavljevic, D., Pekmezovic, T., Milicevic, O., Lazovic, J.M., Cirkovic, A., Savic, M., Rajovic, N., Piperac, P., Djuric, N., Madzarevic, P., Dimitrijevic, A., Randjelovic, S., Nestorovic, E., Akinyombo, R., Pavlovic, A., Ghamrawi, R., Garovic, V., & Milic, N. (2022). How accurate are citations of frequently cited papers in biomedical literature?. *Clinical Science*, *35*(5), 671-681. https://doi.org/10.1042/CS20201573

Peters, M.D.J., Marnie, C., Colquhoun, H., Garritty, C.M., Hempel, S., Horsley, T., Langlois, E.V., Lillie, E., O'Brien, K.K., Tunçalp, Ö, Wilson, M.G., Zarin, W., & Tricco, A.C. (2021). Scoping reviews: reinforcing and advancing the methodology and application. *Systematic Reviews*, *10*, article 263. https://doi.org/10.1186/s13643-021-01821-3

Scheirer, M.A., & Dearing, J.W. (2011). An agenda for research on the sustainability of public health programs. *American Journal of Public Health*, *101*(11), 2059-67. https://doi.org/10.2105/AJPH.2011.300193

Smith, N., & Cumberledge, A. (2020). Quotation errors in general science journals. *Proceedings of the Royal Society A*, 476(2242, 20200538. https://doi.org/10.1098/rspa.2020.0538

Spasojević Brkić, V., Tomić, B., Perišić, M., & Brkić, A. (2020). Kaizen implementation context and performance. *Journal of Engineering Management and Competitiveness*, *10*(1), 31–37. https://doi.org/10.5937/jemc2001031S

Trueblood, J.S., Allison, D.B., Field, S.M., Fishbach, A., Gaillard, S.D.M., Gigerenzer, G., Holmes, W.R., Lewandowsky, S., Matzke, D., Murphy, M.C., Musslick, S., Popov, V., Roskies, A.L., Schure, J.T., & Teodorescu, A.R. (2025). The misalignment of incentives in academic publishing and implications for journal reform. *Proceedings of the National Academy of Sciences of the United States of*

ONLINE FIRST

America, 122(5, article e2401231121. https://doi.org/10.1073/pnas.2401231121

Vanhengel, S., George, B., Letens, G., & Verweire, K. (2026). The organizational change capability of public organizations: Concept and measurement. *Public Money & Management*, *46*(1), 112-123. https://doi.org/10.1080/09540962.2025.250673

Wetzel, R., & Dievernich, F.E.P. (2014). Mind the Gap. The Relevance of Postchange Periods for Organizational Sensemaking. *Systems Research and Behavioral Science*, *31*(2), 280-300. https://doi.org/10.1002/sres.2198

White, A., Smets, M., Tullett, D., Nickerson, S., Canwell, A., Glindemann, C., Fealy, L., Wheelock, M., & Rubenstein, R. (2022), *The future of transformation is human*, Said Business School & EY.

IZA MITA O 70%: ŠTA ZAISTA ZNAMO O STOPAMA NEUSPEHA U INICIJATIVAMA UNAPREĐENJA I TRANSFORMACIJE?

Ovaj rad ispituje šta je zaista poznato o stopama neuspeha u inicijativama unapređenja i transformacije, kao i razloge zbog kojih se tvrdnja o „približno 70% neuspeha“ prenosi kroz različite oblasti. Integrativna sinteza dokaza mapira navode o stopama neuspeha u akademskoj i menadžerskoj literaturi. Svaki navod se kodira prema tipu neuspeha, jedinici analize, imenitelju, vremenskom horizontu i obliku stope, te se procenjuje u pogledu evidencijske pouzdanosti. Prikazane stope znatno variraju i nisu uporedive ukoliko nisu jasno definisani pojmovi, imenitelji i vremenski horizonti. Veliki deo akademskog ponavljanja tvrdnje o „približno 70%“ zasniva se na sekundarnom citiranju ili nedovoljno preciziranim navodima, a ne na ponovnom izračunavanju stopa. Menadžerski pokazatelji često definišu „neuspeh“ kao suprotnost strogo postavljenim kriterijumima potpunog uspeha i/ili kao neodržavanje rezultata tokom višegodišnjih perioda, što dovodi do veće prividne učestalosti neuspeha u poređenju sa užim krajnjim ishodima. U celini posmatrano, dostupni dokazi ne podržavaju tretiranje „70% neuspeha“ kao univerzalnog planskog parametra. Učestalost neuspeha je primerenije razumeti kao raspodelu koja zavisi od toga šta se meri i tokom kog vremenskog perioda. Ova studija redefiniše „debatu o 70%“ kao problem merenja i izveštavanja, integriše akademske i menadžerske dokaze pomoću taksonomije tipova neuspeha i procene njihove pouzdanosti, te predlaže „minimum za izveštavanje o stopama neuspeha“ radi veće proverljivosti i praktične upotrebljivosti budućih navoda.

Ključne reči: Stopa neuspeha; Organizaciona promena; Transformacija; Kontinuirano unapređenje; Merenje.

ONLINE FIRST