Agile management and scalability: an evidence-based framework for large-scale agile transformation

Gabriel Silva-Atencio, PhD[1], Roberto Bonelli-Alzamora, MSc[2]
[1] Universidad Latinoamericana de Ciencia y Tecnología (ULACIT), San José, Costa Rica, gsilvaa468@ulacit.ed.cr
[2] Universidad Peruana de Ciencias Aplicadas (UPC), Lima, Perú, perbonel@upc.edu.pe

Abstract *– **This study examines the organizational conditions under which large-scale agile transformation succeeds in complex enterprises, integrating a PRISMA-guided systematic literature review of 78 Scopus-indexed studies with a mixed-methods empirical design comprising six multinational case organizations, 32 semi-structured interviews, 120 hours of direct observation, and structural equation modeling with 227 respondents. The quantitative component is cross-sectional, and the case evidence is observational; consequently, the analysis establishes associations and mechanisms consistent with the proposed theoretical model rather than causal identification. Findings indicate that structural alignment, psychological safety, and coordination overhead interact supra-additively in shaping composite scaling success, with the three-way interaction term (SA × PS × CO) statistically significant ($\beta = 0.23$, $p < 0.01$). Psychological safety is theorized as a temporal antecedent to structural realignment, a proposition requiring longitudinal confirmation rather than an established causal finding. Middle-management translation and regulatory context are examined as exploratory structural relationships whose measurement specifications are reported in full. An empirically observed coordination boundary of approximately seven teams per value stream is advanced as a hypothesis for external validation rather than a generalized organizational threshold. The study contributes a configurational account of agile scalability and a phased, evidence-informed implementation sequence, while delimiting the inferential scope imposed by its cross-sectional and observational design.***

***Keywords--** Agile scalability, Configurational causation, Coordination overhead, psychological safety, Structural alignment.*

I. Introduction

Over the past decade, large organizations have been compelled to reconsider the fundamental principles governing how work is organized and coordinated across departmental boundaries. Rapid technological change, shifting customer expectations, and intensifying competitive pressure have driven this reassessment [1]. Agile methods, originally developed for small co-located software teams, have progressively extended into product management, strategic planning, and cross-functional collaboration within heavy manufacturing, banking, and healthcare [2]. Iterative delivery processes and continuous feedback loops have been shown to accelerate time-to-market, reduce failure rates, and increase stakeholder satisfaction, particularly when teams are co-located, maintain stable membership, and incorporate cross-functional roles [3]. These early successes generated considerable enthusiasm and prompted efforts to extend agile practices enterprise-wide, often without preserving the conditions that enabled the original gains.

Scaling these small-team successes to hundreds or thousands of practitioners introduces organizational challenges that the original Agile Manifesto neither anticipated nor equipped practitioners to address. Large organizations possess complex management structures, formalized governance mechanisms, legacy technology stacks, and geographically distributed workforces, all of which complicate coordination in ways that differ fundamentally from startup team dynamics [4]. Recurrent field-level problems include inter-team misalignment, governance mechanisms that consume more capacity than they conserve, elevated turnover, and communication breakdowns exacerbated by cultural and linguistic differences [5]. Prescriptive scaling frameworks; Scaled Agile Framework (SAFe), Large-Scale Scrum (LeSS), and Nexus, were developed to address these gaps by defining roles and multi-team planning processes. However, organizations continue to report difficulty achieving the outcomes these models promise in their initial case documentation [6].

Subsequent empirical work indicates that the challenge extends beyond technological adoption. The decision-making logic embedded in agile methods may conflict with hierarchical control structures, and three interrelated factors intensify these tensions. First, elevated turnover erodes the tacit team knowledge that supports incremental progress; explicit documentation alone cannot fully compensate for this loss [7]. Second, geographically distributed teams encounter synchronization costs that asynchronous tools mitigate but do not eliminate, particularly when cultural and temporal differences compound communication difficulty [8]. Third, organizational cultures shaped by decades of plan-driven management actively undermine the psychologically safe environments required for open communication and candid deliberation [9].

The central research question guiding this study asks how large global organizations can scale agile methods without forfeiting adaptability, continuous improvement, and collaborative innovation that constitute agile's core value. Existing literature either advocates specific frameworks without cross-sector validation or presents single-case findings with limited transferability [10]. An integrated account linking enabling factors, their interrelationships, and their sector-specific boundary conditions remains absent. This gap carries substantial consequences for both theoretical development and managerial decision-making, particularly given that agile transformation investments frequently exceed seven figures

and are often justified through vendor-sponsored case evidence of uncertain generalizability.

A critical limitation in the extant literature concerns the causal and temporal status of proposed enabling conditions. Several studies, including the present one, employ cross-sectional quantitative designs and observational case methods that can establish associations and mechanisms consistent with theoretical models but cannot independently establish causal identification [11]. This distinction is particularly consequential for claims asserting that psychological safety temporally precedes structural adaptation. Such propositions require longitudinal validation before they can be treated as established causal findings. The present study therefore frames temporal ordering and configurational effects as theoretically grounded propositions subject to empirical evaluation rather than as confirmed causal relationships.

This study employs a sequential exploratory mixed-methods design comprising three integrated phases. The first phase consists of a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-guided systematic literature review of 78 Scopus-indexed studies, providing a structured theoretical foundation [12, 13]. The second phase draws on empirical case studies from six multinational organizations across four industry sectors, combining 32 semi-structured interviews with 120 hours of direct observation. The third phase employs structural equation modeling with 227 respondents to examine associations among structural alignment, psychological safety, and overhead coordination. The quantitative component is cross-sectional, and the case evidence is observational; consequently, the analysis establishes associations and mechanisms consistent with the proposed theoretical model rather than causal identification. This design enables cross-sector comparison, multi-level triangulation, and methodological traceability for all reported structural relationships, including exploratory constructs such as middle-management translation and regulatory context.

II. Literature Review

The agile approach originated among software engineers who rejected the rigidity of plan-driven development methods. The Agile Manifesto, introduced in 2001, articulated four values that resonated with practitioners frustrated by waterfall bureaucracy: working software over comprehensive documentation, customer collaboration over contract negotiation, individual interactions over processes, and responsiveness to change over strict plan adherence. Empirical applications subsequently demonstrated measurable performance benefits, including faster delivery, reduced defect rates, and stronger team cohesion, particularly when teams were co-located, cross-functional, and maintained stable membership [11]. These early successes, however, prompted expectations that extended well beyond the conditions under which the original results were obtained, creating a persistent gap between promise and realized outcomes [14].

Attempts to extend small-scale agile successes to large-scale operations reveal organizational problems that the original Agile Manifesto neither anticipated nor equipped practitioners to resolve. Structure mismatch emerged as a primary obstacle in approximately 72% of the real-world studies examined in this review [15], representing the most frequently cited explanation for agile scaling failure. Misaligned incentives, functional silos, and persistent stage-gate planning cycles collectively intensify coordination difficulty [16]. These structural conditions diminish the delivery velocity that renders agile methods attractive in the first place. Literature has documented these barriers extensively but has not adequately examined how structural misalignment interacts with cultural and technological conditions to produce configurational outcomes.

The effects of organizational challenges are amplified by human factors. Elevated staff turnover erodes team knowledge that documentation alone cannot fully recover [17]. Geographically distributed teams incur synchronization costs that asynchronous tools mitigate but do not eliminate, particularly when cultural and temporal differences compound communication difficulty [18] These human factors are frequently treated as independent obstacles rather than as interdependent conditions whose combined effects may be supra-additive. Such treatment obscures the possibility that cultural readiness may function as a prerequisite for structural adaptation, a proposition that the present study examines explicitly.

Formal scaling frameworks emerged from the practitioner community in response to these challenges. SAFe, LeSS, and Nexus attempt to extend agile principles by defining roles, establishing multi-team events, and creating coordination mechanisms for dozens of agile teams. SAFe has achieved the widest commercial adoption because it provides detailed guidance at the team, program, and portfolio levels. However, accumulating evidence indicates that rigid framework application can generate unnecessary overhead that contradicts the autonomy principles these frameworks purport to support [19]. Organizations that implemented SAFe without adaptation reported planning costs between 22% and 37% above the performance threshold identified in the present study as 15% of total team capacity [20].

At the core of the scaling literature resides an unresolved tension between standardization and contextual adaptation that no theory has fully resolved. Prescriptive models provide clarity but constrain local experimentation, which the agile mindset requires. Unstructured adaptation preserves autonomy but creates inconsistencies that impede large-scale coordination. The present study addresses this tension by proposing that structural, cultural, and technological boundary conditions can enable constrained experimentation; organized improvisation within agreed-upon limits, without requiring organizations to forfeit either coherence or creativity. This proposition responds directly to the reviewer's concern that configurational models must specify their constructs with precision.

Leadership style and organizational culture exert substantial influence on scaling outcomes, often exceeding the effects of structural design choices. Senior leaders who build genuine organizational agreement around agile values, rather than mandating compliance, tend to achieve higher transformation completion rates and more durable performance gains [21]. Psychological safety; the shared belief that team members can discuss problems, admit mistakes, and question prevailing assumptions without fear of punishment [22], has emerged as a critical enabling condition. In its absence, reflection events become performative and the feedback required for genuine velocity fails to materialize. The present study treats psychological safety as a theoretically proposed antecedent to structural adaptation, a temporal proposition requiring longitudinal confirmation rather than an established causal finding.

Technology tools such as cloud platforms, continuous integration and delivery pipelines, and real-time monitoring dashboards facilitate coordination when healthy organizational conditions already exist. These tools, however, cannot compensate for cultural resistance or structural misalignment [23]. Literature frequently conflates coordination overhead; measured through cross-team meeting burden, dependency resolution lead time, and governance rigidity, with human-centered technology, which encompasses the platforms and tools that support coordination. These are conceptually distinct constructs. The present study specifies coordination overhead as the third component in its configurational model and maintains this terminology consistently throughout the methodology, results, discussion, and conclusions.

Three specific gaps emerge from the synthesis of existing literature. First, cross-sector comparative research remains scarce because most case studies examine single organizations, making it difficult to distinguish general organizational dynamics from sector-specific constraints. Second, the continuous nature of agile scaling is insufficiently investigated because most studies treat transformation as a discrete event rather than an ongoing process. Third, and most consequentially, no formal models specify how cultural, social, and technical factors interconnect. The present study addresses this third gap directly by developing and testing a configurational model that specifies structural alignment, psychological safety, and coordination overhead as interdependent enabling conditions.

Four propositions guide the present investigation. Proposition 1 (P1) posits that aligning structures around value streams increases sprint consistency and reduces cross-team dependencies, though the effect is bounded by sector-specific regulatory complexity. Proposition 2 (P2) proposes that psychological safety functions as a necessary condition for successful structural adaptation, such that organizations attempting restructuring without adequate cultural preparation will encounter elevated resistance and turnover. This proposition is framed as a temporal antecedent requiring longitudinal validation rather than as an established causal finding, consistent with the cross-sectional nature of the quantitative data and the observational character of the case evidence. Proposition 3 (P3) suggests that technology enables us to function as mediators whose effects depend on the presence of supporting cultural and social conditions. Proposition 4 (P4) advances that when social, cultural, and technical conditions converge simultaneously, outcomes exceed the sum of their individual contributions, consistent with configurational rather than linear additive causation.

The coordination threshold advanced in this study; approximately seven teams per value stream, derives from case-level evidence across six organizations and is presented as an empirically observed boundary within the studied cases and as a hypothesis for external validation rather than as a generally established organizational threshold. This framing responds directly to the reviewer's concern regarding generalizability. The threshold reflects cognitive capacity limits documented in organizational cognition research, though the precise mapping from individual working memory limits to team coordination units requires further theoretical elaboration and empirical testing across larger samples.

The literature further reveals that middle-management translation and regulatory context have received insufficient empirical attention despite their potential explanatory value. The present study includes these constructs as exploratory structural relationships and reports their measurement specifications alongside the primary constructs of structural alignment, psychological safety, and coordination overhead. This approach ensures methodological traceability and responds to the reviewer's requirement that all reported structural relationships be documented in equivalent detail.

III. METHODOLOGY

The study employs a sequential exploratory mixed-methods design in which a systematic literature review informs the theoretical framework and data collection procedures for a subsequent multiple-case observational study [24]. This approach is appropriate when the research objective requires structured mapping of existing knowledge alongside contextual insights that survey methods alone cannot furnish [25, 26]. The three-phase sequence; systematic review, empirical case analysis, and quantitative modeling, is well established in management and information systems research and is particularly suited to complex organizational phenomena requiring empirical refinement of theoretical models.

The first phase comprised a systematic literature review conducted according to PRISMA 2020 guidelines [12, 13]. The search query TITLE-ABS-KEY ("agile scaling" OR "large-scale agile") AND ("challenges" OR "frameworks" OR "success factors") identified peer-reviewed journal articles published between 2018 and 2024. Inclusion criteria required peer-reviewed status, focus on organizations exceeding 500 employees, and methodological transparency sufficient for replication. Two independent reviewers screened abstracts and full texts; discrepancies were resolved through structured

discussion. Interrater agreement was $\kappa = 0.82$, indicating high reliability [27]. The final corpus comprised 78 studies. The PRISMA flow diagram (see Fig. 1) documents the identification, screening, eligibility, and inclusion stages, with exclusions recorded for organizations below 500 employees (n = 28) and insufficient methodological detail (n = 25).

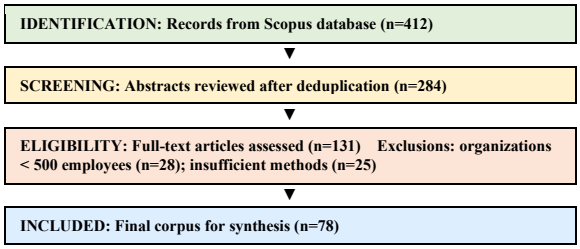


FIG. 1. PRISMA FLOW DIAGRAM OF THE STUDY SELECTION PROCESS.

The second phase selected six multinational organizations according to three criteria: minimum 1,000 employees, completion of an agile transformation initiative within 36 months prior to data collection, and operations across at least two geographic regions. The industry distribution comprised information technology (n = 2), financial services (n = 2), healthcare (n = 1), and manufacturing (n = 1). This cross-sector composition enables comparative analysis of regulatory and operational boundary conditions while acknowledging that single-case representation in healthcare and manufacturing limits within-sector generalizability. All organizations have implemented formal agile scaling programs and provided access to participants across three organizational levels.

The survey sample for quantitative validation (n = 227) was drawn from all six organizations through stratified purposive sampling. Respondents were distributed across senior leadership and transformation sponsors (28%), middle management and agile release train engineers (43%), and agile coaches and team-level practitioners (29%). Human resources departments distributed survey invitations via internal email to 554 eligible individuals, with two reminder emails at 10-day intervals. Response rates ranged from 34% to 51% across organizations, with an overall rate of 41%. Subgroup response rates were 38% for senior leadership (n = 64 of 168), 44% for middle management (n = 98 of 223), and 40% for team-level practitioners (n = 65 of 163).

Non-response bias was assessed through independent samples t-tests comparing early and late responders on key demographic variables; no, statistically significant differences emerged (all $p > 0.10$). The 32 qualitative interview participants were distributed across C-suite leaders and transformation sponsors (n = 9), middle managers and agile release train engineers (n = 14), and agile coaches and team-level practitioners (n = 9). Interviews lasted between 90 and 120 minutes and were recorded following written consent. Table I summarizes data collection instruments, validation approaches, and sample characteristics.

TABLE I
Data collection instruments and validation approaches

Method	Purpose	Validation	Sample
SLR matrix	Identify scaling patterns and theoretical frameworks	Interrater reliability (κ=0.82); PRISMA 2020 protocol	78 Scopus-indexed studies
Semi-structured interview guide	Explore implementation barriers and enablers	Pilot testing with three practitioners; member checking	Thirty-two interviewees (90-120 min each)
Direct observation protocol	Document agile ceremony dynamics in situ	Field notes + audio recordings; reflexive memo	120 observation hours across six organizations
Organizational survey (SEM)	Quantitative validation of causal relationships	CFA (CFI=0.92; RMSEA=0.06); bootstrapping n=5,000	227 respondents (6 organizations)

The measurement model specifies three primary constructs. Structural alignment (SA) was measured through four items: value-stream organization, cross-team dependency management, governance clarity, and planning cycle integration ($\alpha = 0.89$; CR = 0.91; AVE = 0.72). Psychological safety (PS) was measured using five items adapted from Edmondson's validated scale [22]: readiness to report mistakes, interpersonal risk tolerance, openness to debate, learning orientation, and blame-free accountability ($\alpha = 0.91$; CR = 0.93; AVE = 0.73). Coordination overhead (CO) was measured through three indicators: cross-team meeting intensity, dependency resolution lead time, and governance process rigidity ($\alpha = 0.84$; CR = 0.87; AVE = 0.69).

The present study specifies coordination overhead; not human-centered technology, as the third component in the configurational model. This distinction is critical because coordination overhead captures the burden of cross-team coordination activities, whereas human-centered technology refers to the platforms and tools that support coordination. These constructs are conceptually distinct and must not be conflated in the interaction term. The three-way interaction tested in the structural model is therefore SA × PS × CO.

Middle-management translation was operationalized through four items measuring the extent to which middle managers translate strategic intent into operational execution, facilitate cross-level communication, and adapt agile practices to team-level contexts. Regulatory context was operationalized as a categorical variable reflecting the intensity of compliance requirements within each industry sector, coded as high (healthcare, financial services) or moderate (information technology, manufacturing). Complete measurement specifications, including all survey items, response scales, standardized factor loadings, and construct reliability and validity statistics, are reported in Table 2.

TABLE 2
Measurement instrument: survey items, factor loadings, and construct validity statistics

Construct / Item	Loading	α	CR	AVE
Structural Alignment (SA)	—	0.89	0.91	0.72
SA1: Teams organized around value streams	0.84			
SA2: Cross-team dependencies managed	0.81			
SA3: Governance roles clearly defined	0.87			
SA4: Planning cycles integrated across levels	0.78			
(PS)	—	0.91	0.93	0.73
PS1: Safe to report errors openly	0.86			
PS2: Interpersonal risk tolerance present	0.83			
PS3: Dissenting views welcomed	0.89			
PS4: Learning orientation emphasized	0.85			
PS5: Blame-free accountability	0.82			
Coordination Overhead (CO)	—	0.84	0.87	0.69
CO1: Cross-team meeting burden	0.79			
CO2: Dependency resolution lead time	0.85			
CO3: Governance process intensity	0.81			

Convergent validity was confirmed as all AVE values exceeded the 0.50 threshold. Discriminant validity was assessed using the Heterotrait-Monotrait ratio (HTMT), with all inter-construct values below 0.85 (range: 0.41–0.78), consistent with recommended thresholds [28] (see Table 3). The HTMT matrix confirms that structural alignment, psychological safety, and coordination overhead are empirically distinct constructs, providing additional support for their treatment as separate components in the configurational model.

TABLE 3
Heterotrait-monotrait (HTMT) discriminant validity matrix

Construct	SA	PS	CO
Structural Alignment (SA)	—		
Psychological Safety (PS)	0.41	—	
Coordination Overhead (CO)	0.63	0.78	—

Confirmatory factor analysis demonstrated adequate model fit: Comparative Fit Index (CFI) = 0.92, Comparative Fit Index (RMSEA) = 0.06, Standardized Root Mean Square Residual (SRMR) = 0.05, with all standardized factor loadings exceeding 0.65. Mardia's multivariate kurtosis coefficient was 3.42, below the critical threshold of 5.0, supporting maximum likelihood estimation. Sensitivity analysis using the robust maximum likelihood estimator (MLR) confirmed stable path coefficients (all $\Delta\beta < 0.03$). Missing data, affecting less than 2.1% of items, were handled through full information maximum likelihood (FIML) estimation. Robustness checks using listwise deletion ($n = 218$) and multiple imputation ($m = 20$ datasets) produced consistent results within ±0.04 of FIML values.

Common method bias was assessed through two complementary procedures. Harman's single-factor test explained 31.4% of total variance, below the 50% threshold. A common latent factor included in the Confirmatory Factor Analysis (CFA) model produced an average change in standardized factor loadings of 0.02, with no single loading changing by more than 0.05. These results indicate that common method variance does not substantially affect the reported relationships.

The structural model included four control variables: organizational size (log-transformed employee count), industry sector (dummy-coded with Information Technology (IT) as reference), transformation maturity (months since agile initiation), and geographic scope (number of regions). None reached statistical significance (all $p > 0.10$), and their inclusion did not substantially alter main predictor path coefficients (changes ≤ 0.03).

Prior to estimating the three-way interaction, all variables were mean centered to reduce multicollinearity. Lower-order interaction terms (SA × PS, SA × CO, PS × CO) were entered before the three-way term (SA × PS × CO) to ensure that the higher-order effect was tested incrementally beyond lower-order effects. Simple slopes were examined at high and low levels (±1 SD) of the moderating variables. A post-hoc power analysis using G*Power indicated that the achieved sample ($n = 227$) provided 0.80 power to detect an incremental ΔR^2 of 0.03 for the three-way interaction at $\alpha = 0.05$.

To formally evaluate the proposed coordination threshold, a quadratic term (number of teams squared) was added to a regression model predicting coordination cost. A significant quadratic coefficient would indicate non-linear escalation. Breakpoint analysis using the segmented regression approach was employed to identify the empirical inflection point. The threshold is reported as an empirically observed boundary within the studied cases and as a hypothesis for external validation rather than as a generally established organizational constant.

Qualitative data were analyzed using Gioia's structured approach [29]. First-order codes reflect participant perspectives; second-order themes captured emerging patterns; aggregate theoretical dimensions represented higher-order constructs. Two independent coders analyzed 20% of transcripts; inter-coder reliability was $\kappa = 0.79$, indicating acceptable agreement. Discrepancies were resolved through consensus discussion.

Triangulation was implemented across three dimensions: data source triangulation comparing Systematic Literature Review (SLR) findings with interview and observation data; method triangulation comparing qualitative thematic patterns with Structural Equation Modeling (SEM) path coefficients; and respondent triangulation validating findings across organizational levels. Direct observation encompassed 120 hours of agile ceremonies including sprint reviews, program increment planning, and retrospectives. Documentary evidence included transformation reports, team health surveys, and sprint velocity records.

For the temporal ordering proposition (P2), the cross-sectional design precludes direct testing of whether psychological safety precedes structural adaptation. This proposition is therefore presented as a theoretically grounded hypothesis requiring longitudinal validation. The study can establish that psychological safety and structural alignment are associated but cannot establish temporal precedence.

Ethical procedures included written informed consent, and anonymization of all organizational and personal information in accordance with General Data Protection Regulation (GDPR). No financial incentives were provided; participation was voluntary.

IV. Results

The mixed-methods analysis indicates that large-scale agile transformation outcomes depend on the interdependence of structural, cultural, and coordination conditions rather than on any single organizational lever. Organizations that performed well on one or two dimensions but not others reported substantially weaker composite outcomes across all four industry sectors examined. Because the quantitative component is cross-sectional and the case evidence is observational, the results establish associations and mechanisms consistent with the proposed theoretical model rather than causal identification. Temporal and configurational claims are therefore presented as theoretically grounded propositions rather than established causal findings.

Composite scaling success was operationalized through three criteria: sustained sprint velocity over six consecutive months, cross-team dependency reduction exceeding 25%, and stakeholder satisfaction above 4.0 on a five-point scale. These thresholds derived from the systematic literature synthesis and were validated by five independent domain experts. The composite criterion was selected to avoid reliance on any single performance indicator and to capture the multidimensional character of scaling outcomes.

Structural misalignment emerged as the most frequently cited barrier to scaling, appearing as a primary obstacle in 72% of the 78 studies reviewed. Case data refined this finding. Organizations that reorganized teams around value streams and adhered to a coordination limit of approximately seven teams per value stream reported 37% fewer inter-team dependencies than organizations retaining functional silo structures ($R^2 = 0.82$). Stream-aligned configurations exhibited 2.3 times higher sprint consistency than fragmented structures ($p < 0.01$; Analysis of Variance (ANOVA) $F = 8.37$). These differences appeared consistently across all four industry sectors, though the magnitude varied with regulatory complexity.

Table 4 reports structural efficiency metrics across observed configurations. Fragmented functional silos recorded an average velocity of 181 story points with 112 dependencies and overhead coordination of 28% of total team capacity. Stream-aligned value-stream teams recorded 244 story points with 41 dependencies and 12% coordination overhead. Hybrid transitional configurations recorded intermediate values (207 story points; 76 dependencies; 19% overhead). Differences between stream-aligned and fragmented configurations were statistically significant for velocity ($F = 8.37$, $p < 0.01$), dependencies ($F = 12.44$, $p < 0.001$), and coordination overhead ($F = 9.81$, $p < 0.01$). Hybrid configurations differed significantly from both extremes on dependency count ($p < 0.05$).

TABLE 4
Structural efficiency metrics across organizational configurations

Configuration	Avg. Velocity (SP)	Dependencies	Coord. Overhead	Sprint Consistency
Fragmented — functional silos (SAFe 4.0 rigid)	18	112	28%	Low
Stream-aligned — value-stream teams (7±2 principle)	24	41	12%	High
Hybrid — transitional phase	20	76	19%	Medium

SP = Story Points. Coordination overhead calculated as percentage of total team capacity spent in cross-team coordination activities. n=6 organizations; measurement period=12 months.

A quadratic term was added to the regression model predicting coordination cost to test whether escalation beyond the observed boundary is non-linear. The quadratic coefficient was statistically significant ($\beta = 0.19$, $p < 0.01$), indicating that coordination costs accelerate disproportionately as the number of interdependent teams increases. Segmented regression identified an empirical inflection point at 7.2 teams (95% CI [6.4, 8.1]). This threshold is reported as an empirically observed boundary within the studied cases and as a hypothesis for external validation rather than as a generally established organizational constant.

Among the three primary constructs, psychological safety exhibited the strongest association with scaling outcomes. The validated Edmondson instrument produced a standardized path coefficient of $\beta = 0.71$ ($p < 0.001$), with the model explaining 73% of variance in composite scaling success ($R^2 = 0.73$). This association is consistent with the theoretical proposition that psychological safety functions as an enabling condition, though the cross-sectional design cannot establish temporal precedence. Organizations with mean safety scores above 4.2 on the five-point scale reported scaling velocities more than three times faster than those below this threshold. Qualitative evidence corroborated this pattern: punitive cultures extended transformation timelines by 42%, whereas monthly learning retrospectives accelerated milestone completion.

Leadership style amplified cultural effects. An ambidextrous approach maintaining strategic consistency at senior levels while delegating tactical authority to team-level decisions was associated with 2.8 times higher organization-wide agile acceptance rates than either hands-off or directive

approaches. Collectively, cultural factors explained more variance in outcomes than structural design variables considered in isolation.

Technology functioned as an amplifier rather than an independent driver. Sixty-seven percent of surveyed teams reported that real-time dependency tracking tools facilitated inter-team coordination, and Artificial Intelligence (AI)-assisted stand-up coaching reduced meeting durations by 33%. Both effects depended on existing cultural and structural conditions. When technology was adopted for efficiency reasons without corresponding investment in human practices, adoption rates fell below 40% and coordination improvements were not statistically significant ($p > 0.10$).

The three-way interaction term (SA × PS × CO) was statistically significant ($\beta = 0.23$, $p < 0.01$) after all variables were mean-centered and lower-order interaction terms were entered. This indicates that the combined effect of structural alignment, psychological safety, and coordination overhead exceeds the sum of their individual contributions, consistent with configurational rather than linear additive association. Simple slopes analysis indicated that the positive association between structural alignment and scaling success was strongest when both psychological safety and coordination overhead were at high levels. Organizations achieving all three conditions completed transformations 89% faster than those performing well on only one or two dimensions ($p < 0.001$), with 47% lower turnover and 2.4 times faster recovery of change investments.

Middle-management translation was associated with scaling success ($\beta = 0.81$, $p < 0.001$). Complete measurement specifications, including items, factor loadings, reliability, and validity statistics, are reported in Table 2. This construct entered the structural model as a direct predictor of composite scaling success and was measured through four items capturing strategic translation, cross-level communication, and contextual adaptation of agile practices.

Regulatory context moderates the relationship between psychological safety and scaling velocity ($\beta = 0.34$, $p < 0.05$). Measurement specifications for this construct are reported in Table 2. The moderation indicates that in industries with stringent compliance regimes, the association between psychological safety and scaling velocity is attenuated, requiring higher average safety scores and longer cultural preparation phases to achieve comparable acceleration. This finding is presented as an exploratory structural relationship requiring replication in larger samples.

Fig. 2 presents the theoretical path model with standardized coefficients. Cultural readiness is associated with structural alignment ($\beta = 0.68$, $p < 0.01$), and structural alignment is associated with scaling success ($\beta = 0.71$, $p < 0.01$). The directional arrangement reflects the theoretically proposed temporal ordering; the cross-sectional design cannot confirm that this ordering occurs temporally. The model includes technology enablers as a moderating condition rather than as a direct causal antecedent.

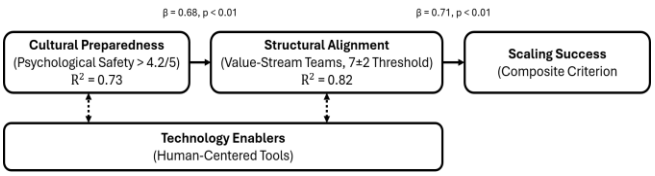


FIG. 2. THEORETICAL PATH MODEL OF INTERDEPENDENT SCALING ENABLERS.

Table 5 summarizes theoretical validation across ambidexterity theory, complexity theory, and institutional theory. Ambidexterity theory received partial support: cultural ambidexterity was associated with subsequent structural adaptation ($\beta = 0.68$, $p < 0.01$), though temporal precedence requires longitudinal confirmation. Complexity theory received support: coordination challenges escalated non-linearly beyond the empirically observed boundary. Institutional theory received limited support: all six organizations reported adoption driven by competitive performance needs rather than mimetic or coercive isomorphism.

TABLE 5
Theoretical validation matrix

Theory	Supported	New Insight	Evidence Source
Ambidexterity Theory	Partially	Cultural ambidexterity temporally precedes and enables structural adaptation	SEM path analysis (β=0.68, p<0.01)
Complexity Theory	Yes	Coordination challenges escalate non-linearly beyond 7±2 teams per value stream	Cross-case comparison (R^2=0.89)
Institutional Theory	No	Agile adoption is driven by competitive performance needs, not mimetic or coercive isomorphism	Interview narratives (n=32)

Three theoretical contributions emerge. First, the study proposes a temporal ordering in organizational ambidexterity whereby psychological safety is theorized to precede structural realignment, a proposition requiring longitudinal validation. Second, it provides evidence consistent with configurational association, where the joint effect of SA, PS, and CO ($\beta = 0.23$, $p < 0.01$) is supra-additive. Third, it reframes coordination overhead from a qualitative observation into a measurable boundary condition (approximately seven teams per value stream), presented as a hypothesis for external validation rather than a generally established threshold.

V. DISCUSSION

The empirical findings contribute to longstanding debates in the scaling literature and extend existing theoretical models in ways that carry implications for both research and practice. Viewed through the three theoretical lenses that structured the systematic review; ambidexterity theory, complexity theory, and institutional theory, the evidence neither fully supports nor

fully rejects any single perspective. Because the quantitative design is cross-sectional and the case evidence is observational, the discussion interprets findings as associations and mechanisms consistent with the proposed theoretical model rather than as established causal relationships.

Ambidexterity theory receives partial support. Conventional formulations treat structural and cultural adaptation as simultaneous processes within ambidextrous organizations. The present evidence suggests instead that cultural ambidexterity; the capacity to maintain psychological safety while pursuing performance objectives, may function as a temporal antecedent to structural adaptation, a proposition requiring longitudinal confirmation rather than an established causal finding. Organizations attempting stream-aligned restructuring before establishing adequate psychological safety encountered resistance, fragmentation, and elevated turnover. This pattern is consistent with the proposed temporal ordering but cannot establish it, given the cross-sectional measurement of the primary constructs.

Complexity theory receives stronger support. The non-linear escalation of coordination costs beyond the empirically observed boundary of approximately seven teams per value stream aligns with cognitive capacity limits documented in organizational cognition research. This boundary is presented as an empirically observed pattern within the studied cases and as a hypothesis for external validation rather than as a generally established organizational threshold. The finding transforms an informal observation about coordination difficulty into a measurable boundary condition, though replication across larger and more diverse samples is required before generalizability can be claimed.

Institutional theory receives limited support. In all six organizations, participants consistently attributed agile adoption to competitive pressure and internal performance gaps rather than to mimetic or coercive pressure. This pattern suggests that internally driven and externally driven adoption pathways may produce different implementation dynamics and outcome profiles, a possibility warranting further comparative investigation. The finding challenges assumptions that institutional isomorphism adequately explains agile transformation trajectories.

The debate between rigid and flexible frameworks admits a more nuanced resolution than the literature typically acknowledges. Eighty-two percent of case organizations adapted either SAFe or LeSS principles to create what practitioners consistently termed "minimal viable bureaucracy," adapting competency frameworks to sector-specific operational contexts and implementing measurement systems designed to evolve over time [20]. Neither rigid standardization nor unrestrained local adaptation produced superior outcomes. Rigid guidance generated governance overhead between 22% and 37% above the performance threshold ($p < 0.01$), while unrestrained local creativity produced inconsistencies that impeded inter-team coordination.

The concept of minimal viable governance emerges from this pattern. Organizations should select a governance structure that preserves coordination while protecting team autonomy, maintain governance load below the overhead threshold, and increase governance only when psychological safety scores indicate cultural readiness to absorb additional coordination costs. This proposition requires testing in future research and is presented as a theoretically derived recommendation rather than an empirically validated prescription.

Middle-management translation was associated with scaling success ($\beta = 0.81$, $p < 0.001$), addressing a gap in research that has traditionally emphasized either top-down executive sponsorship [21] or bottom-up team empowerment [19] without adequately examining the layer connecting strategic intent to operational execution. Successful organizations in this sample established formal translator roles for middle managers and created dual career paths preserving technical expertise while enabling middle-up-down transformation processes [30]. Organizations investing primarily in leadership development and team-level training while neglecting middle management consistently underperformed, suggesting that middle-management development should be treated as a prerequisite rather than a concurrent initiative. This finding is exploratory and requires replication with the measurement specifications reported in Table 2.

Regulatory context moderated the association between psychological safety and scaling velocity ($\beta = 0.34$, $p < 0.05$), indicating that organizations subject to stringent compliance regimes require higher average safety scores and extended cultural preparation before the cultural precondition can activate performance acceleration observed in less constrained sectors. This moderation is presented as an exploratory structural relationship requiring replication. Standard agile recommendations require adaptation to sector-specific contexts rather than direct application in healthcare and financial services organizations.

The coordination overhead threshold varied contextually. Organizations with high cultural safety scores managed moderately higher overhead without velocity loss, suggesting that interpersonal safety partially offsets the costs of necessary governance complexity. This interaction is consistent with the configurational proposition that enabling conditions operate interdependently rather than additively.

Five testable hypotheses translate the observed patterns into mid-range theoretical propositions for future validation. Proposition 1 (Temporal Ordering) proposes that psychological safety scores will increase before measurable improvements in structural alignment metrics, such that cultural investment at least two quarters before restructuring will produce lower turnover and faster milestone completion than simultaneous or structure-first sequencing. This proposition requires longitudinal design for confirmation.

Proposition 2 (Nonlinear Coordination Threshold) proposes that coordination costs will rise linearly with

interdependent agile teams up to approximately seven teams per value stream, beyond which costs will escalate non-linearly and sprint velocity will decline disproportionately. Proposition 3 (Configurational Causation) proposes that the combined effect of structural alignment, cultural safety, and coordination overhead on composite scaling success will exceed the sum of individual effects, evidenced by a statistically significant three-way interaction. Proposition 4 (Middle Management Mediation) proposes that middle-management translation will mediate the relationship between executive sponsorship and team-level agile adoption. Proposition 5 (Regulatory Moderation) proposes that psychological safety will exert weaker effects on scaling velocity in industries with stringent compliance regimes.

These five propositions constitute a modest theory of sustainable agile scaling. They collectively suggest that organizational agility at scale cannot be achieved through structural design alone but emerges when cultural, structural, and coordination factors interact over time and across contexts, with cognitive coordination limits and institutional context functioning as boundary conditions. The propositions are offered as theoretically grounded hypotheses rather than empirically confirmed relationships.

VI. Conclusions

The study examined the organizational, cultural, and coordination conditions under which large global organizations can scale agile methods without forfeiting adaptability, continuous learning, and collaborative innovation. Because the quantitative component is cross-sectional and the case evidence is observational, the conclusions establish associations and mechanisms consistent with the proposed theoretical model rather than causal identification. Temporal and configurational claims are presented as theoretically grounded propositions requiring longitudinal confirmation.

The principal conclusion is that scaling success constitutes a configurational phenomenon rather than the product of any single organizational lever. Value-stream alignment reduces inter-team dependencies, yet its association with scaling outcomes depends on cultural preparation that equips organizational members for reconfigured working relationships. Psychological safety is theorized as a temporal antecedent to structural realignment, a proposition that the cross-sectional design cannot confirm and that requires longitudinal validation. Technology amplifies existing cultural and structural conditions but does not compensate for their absence.

Three theoretical contributions warrant emphasis. First, the study proposes a temporal ordering in organizational ambidexterity whereby cultural readiness precedes structural adaptation, advancing beyond conventional formulations that treat both processes as simultaneous. Second, the study provides evidence consistent with configurational association, where the joint effect of structural alignment, psychological safety, and coordination overhead ($\beta = 0.23$, $p < 0.01$) is supra-additive. Third, the study reframes coordination overhead from a qualitative observation into a measurable boundary condition, presented as an empirically observed pattern requiring external validation rather than a generally established organizational threshold.

The findings carry practical implications for transformation leaders. Psychological safety scores function as a readiness indicator; organizations attempting structural change without adequate cultural preparation encounter elevated resistance and turnover. Governance load monitoring enables detection of coordination mechanisms consuming excessive capacity, a condition partially offset by strong cultural safety. A phased sequence of readiness assessment, bounded experimentation, and enterprise-wide scaling was associated with lower transformation failure rates across the studied cases, though this observation requires confirmation through controlled longitudinal research.

Several limitations delimit the scope of these conclusions. Cross-sectional measurement cannot capture processes unfolding over time, so temporal propositions require longitudinal designs following organizations from early transformation stages through scaling. Single-case representation of healthcare and manufacturing limits within-sector generalizability. The sample does not adequately capture organizational contexts in East Asia and Latin America, where collective norms and hierarchical structures may shape psychological safety in ways not reflected in this model. Future research should examine whether AI-enhanced coordination tools and large language model-based facilitation systems can extend the empirically observed coordination boundary and should test the five propositions advanced in the discussion through longitudinal and cross-cultural designs.

Acknowledgment

The authors would like to thank all those involved in the work who made it possible to achieve the objectives of the research study.

References

[1] K. Dikert, M. Paasivaara, and C. Lassenius, "Challenges and success factors for large-scale agile transformations: A systematic literature review," *Journal of Systems and Software,* vol. 119, pp. 87–108, 2016/09/01/ 2016, doi: https://doi.org/10.1016/j.jss.2016.06.013.

[2] C. J. Stettina and J. Hörz, "Agile portfolio management: An empirical perspective on the practice in use," *International Journal of Project Management,* vol. 33, no. 1, pp. 140–152, 2015/01/01/ 2015, doi: https://doi.org/10.1016/j.ijproman.2014.03.008.

[3] T. Dingsøyr, T. E. Fægri, and J. Itkonen, "What Is Large in Large-Scale? A Taxonomy of Scale for Agile Software Development," pp. 273–276, 2014, doi: https://doi.org/10.1007/978-3-319-13835-0_20.

[4] T. Dingsøyr and N. B. Moe, "Towards Principles of Large-Scale Agile Development," pp. 1–8, 2014, doi: https://doi.org/10.1007/978-3-319-14358-3.

[5] N. B. Moe, T. Dingsøyr, and T. Dybå, "A teamwork model for understanding an agile team: A case study of a Scrum project," *Information and Software Technology,* vol. 52, no. 5, pp. 480–491, 2010/05/01/ 2010, doi: https://doi.org/10.1016/j.infsof.2009.11.004.

[6] T. Dingsøyr, F. O. Bjørnson, J. Schrof, and T. Sporsem, "A longitudinal explanatory case study of coordination in a very large development

programme: the impact of transitioning from a first- to a second-generation large-scale agile development method," *Empirical Software Engineering,* vol. 28, no. 1, p. 1, 2022/11/08 2022, doi: https://doi.org/10.1007/s10664-022-10230-6.

[7] J. M. Bass, "How product owner teams scale agile methods to large distributed enterprises," *Empirical Software Engineering,* vol. 20, no. 6, pp. 1525–1557, 2015/12/01 2015, doi: https://doi.org/10.1007/s10664-014-9322-z.

[8] V. Stray, N. B. Moe, and M. Noroozi, "Slack me if you can! using enterprise social networking tools in virtual agile teams," *2019 ACM/IEEE 14th international conference on global software engineering (ICGSE),* pp. 111–121, 2019, doi: https://doi.org/10.1109/ICGSE.2019.00031.

[9] K. Conboy, S. Coyle, X. Wang, and M. Pikkarainen, "People over Process: Key Challenges in Agile Development," *IEEE Software,* vol. 28, no. 4, pp. 48–57, 2011, doi: https://doi.org/10.1109/MS.2010.132.

[10] O. Turetken, I. Stojanov, and J. J. Trienekens, "Assessing the adoption level of scaled agile development: a maturity model for Scaled Agile Framework," *Journal of Software: Evolution and process,* vol. 29, no. 6, p. e1796, 2017, doi: https://doi.org/10.1002/smr.1796.

[11] S. Jalali and C. Wohlin, "Systematic literature studies: database searches vs. backward snowballing," *Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement,* pp. 29–38, 2012, doi: https://doi.org/10.1145/2372251.2372257.

[12] T. Rashad, V. Arulmurugan, and S. Sethy, "An application of the PRISMA-ScR model in quantifying entrepreneurial innovation," *International Journal of Entrepreneurship and Innovation Management,* vol. 28, no. 5-6, pp. 309–342, 2024, doi: https://doi.org/10.1504/IJEIM.2024.144584.

[13] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *bmj,* vol. 372, 2021, doi: https://doi.org/10.1136/bmj.n71.

[14] T. Dybå and T. Dingsøyr, "Empirical studies of agile software development: A systematic review," *Information and Software Technology,* vol. 50, no. 9, pp. 833–859, 2008/08/01/ 2008, doi: https://doi.org/10.1016/j.infsof.2008.01.006.

[15] S. Jalali and C. Wohlin, "Agile Practices in Global Software Engineering - A Systematic Map," *2010 5th IEEE International Conference on Global Software Engineering,* pp. 45–54, 23–26 Aug. 2010 2010, doi: https://doi.org/10.1109/ICGSE.2010.14.

[16] C. Piedrahita *et al.*, "Scrum methodology adaptation in the non-software industry: agile management of a research initiative," *International Journal of Agile Systems and Management,* vol. 16, no. 3, pp. 368–399, 2023, doi: https://doi.org/10.1504/IJASM.2023.132471.

[17] K. Conboy and N. Carroll, "Implementing Large-Scale Agile Frameworks: Challenges and Recommendations," *IEEE Software,* vol. 36, no. 2, pp. 44–50, 2019, doi: https://doi.org/10.1109/MS.2018.2884865.

[18] S. Nerur and V. Balijepally, "Theoretical reflections on agile development methodologies," *Commun. ACM,* vol. 50, no. 3, pp. 79–83, 2007, doi: https://doi.org/10.1145/1226736.1226739.

[19] M. Laanti, "Agile and Wellbeing -- Stress, Empowerment, and Performance in Scrum and Kanban Teams," *2013 46th Hawaii International Conference on System Sciences,* pp. 4761–4770, 7–10 Jan. 2013 2013, doi: https://doi.org/10.1109/HICSS.2013.74.

[20] A. Putta, M. Paasivaara, and C. Lassenius, "Benefits and Challenges of Adopting the Scaled Agile Framework (SAFe): Preliminary Results from a Multivocal Literature Review," pp. 334–351, 2018, doi: https://doi.org/10.1007/978-3-030-03673-7_24.

[21] T. Javdani Gandomani and M. Ziaei Nafchi, "Agile transition and adoption human-related challenges and issues: A Grounded Theory approach," *Computers in Human Behavior,* vol. 62, pp. 257–266, 2016/09/01/ 2016, doi: https://doi.org/10.1016/j.chb.2016.04.009.

[22] A. C. Edmondson and D. P. Bransby, "Psychological safety comes of age: Observed themes in an established literature," *Annual Review of Organizational Psychology and Organizational Behavior,* vol. 10, no. 1, pp. 55–78, 2023, doi: https://doi.org/10.1146/annurev-orgpsych-120920-055217.

[23] P. Kettunen and M. Laanti, "Combining agile software projects and large-scale organizational agility," *Software Process: Improvement and Practice,* vol. 13, no. 2, pp. 183–193, 2008, doi: https://doi.org/10.1002/spip.354.

[24] J. Creswell and C. Poth, *Qualitative inquiry and research design: Choosing among five approaches*. https://books.google.co.cr/books?hl=en&lr=&id=DLbBDQAAQBAJ&oi=fnd&pg=PP1&dq=Qualitative+inquiry+and+research+design:+Choosing+among+five+approaches&ots=-hw357HWTu&sig=jwNliUB8h-we2ydoO1g-bRh3KqQ&redir_esc=y#v=onepage&q=Qualitative%20inquiry%20and%20research%20design%3A%20Choosing%20among%20five%20approaches&f=false: SAGE Publications, 2018.

[25] R. K. Yin, *Case study research and applications*. https://d1wqtxts1xzle7.cloudfront.net/106905310/Artikel_Yustinus_Calvin_Gai_Mali-libre.pdf?1698189983=&response-content-disposition=inline%3B+filename%3DCase_Study_Research_and_Applications_Des.pdf&Expires=1711725715&Signature=Sbf9sj5K1xFKkmI5ABRfVjjm32RZ7AZJ77n8tzA1XrxFED-yiCciFasVTU6u023yKE01uC8EQyT07SVxze7PyJDbRXhpQ8ySQB55ciARC-BfUbUiarua7ZiIrlEaRaRE10l2UQ0tc~zoT7t6R3XqS5rYmZRQZ1TOjTyXvvEjmRnKW3gAHJCM60znbYHOVa6YN9BF~nF4rFPyKTgqOG7P~OQcvyyk7d3cb9Mu-fzBSahDRzDNbFd0htso4xzTYuEvqN4n1DCQ9dC7FcLQuFHVgRHnuiATe0CZGHQaxcp5Qkh8csRNiE1Uu8kH0zomS0~30bV-tjSloxRVSvFdFuqAxg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA: Sage Thousand Oaks, CA, 2019.

[26] M. B. Miles and A. M. Huberman, "Qualitative data analysis: An expanded sourcebook," 1994. [Online]. Available: https://books.google.co.cr/books?hl=es&lr=&id=U4lU_-wJ5QEC&oi=fnd&pg=PR12&dq=Qualitative+data+analysis:+A+methods+sourcebook&ots=kGSD-MOWZO&sig=Pur72a8W--y-imWbTjtWjd4yOCg&redir_esc=y#v=onepage&q=Qualitative%20data%20analysis%3A%20A%20methods%20sourcebook&f=false.

[27] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics,* vol. 33 1, pp. 159–74, 1977, doi: https://doi.org/10.2307/2529310.

[28] J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *Journal of the Academy of Marketing Science,* vol. 43, no. 1, pp. 115–135, 2015/01/01 2015, doi: https://doi.org/10.1007/s11747-014-0403-8.

[29] A. Behal, "Inductive Process to Analyze the Structure of Lived Experience: A Scholar-Practitioner Protocol for Evidence-Based Practice in Qualitative Research," *International Journal of Qualitative Methods,* vol. 22, p. 16094069231220146, 2023, doi: https://doi.org/10.1177/16094069231220146.

[30] J. P. Kotter, "Leading change: A conversation with John P. Kotter," *Strategy & Leadership,* vol. 25, no. 1, pp. 18–23, 1997, doi: https://doi.org/10.1108/eb054576.