Measuring Scientific Decision-Making in Organizations

May 4, 2026

Abstract

This research note introduces a 24-item survey module for measuring reported scientific decision-making (SDM) practices in organizations, reports diagnostic and within-study consistency checks, and applies it to a field sample of 3,178 respondents in 54 firms. Scientific decision-making is increasingly treated as a managerial resource, yet scalable tools to observe *reported* decision practices inside firms remain scarce. Main-text statistics for overall SDM, vignette-linked outcomes, and hierarchy regressions use the headline analytic sample ($N = 2{,}745$) under the item-weighted scoring rule described in Data and Sample. In that sample, variation is concentrated within firms: the overall SDM intraclass correlation is 0.018, implying that only 1.8% of variance is attributable to firm membership. Adaptive Learning, the dimension capturing whether managers revise beliefs in response to new evidence, shows the clearest top-tier contrast: Tier 1 managers score higher than frontline managers (unconditional $\Delta = .17$; with education, tenure, and decision-area controls, $b = .14$–$.16$, $p < .001$, without and with firm fixed effects), while the other dimensions, Causal Reasoning and Evidence Evaluation, do not show comparable top-tier patterns. SDM is also positively associated with performance on a short within-study decision vignette ($r = .19$), with an approximately nine-percentage-point top-versus-bottom-quintile gap on vignette accuracy (9.3 percentage points before rounding). The contribution of this note is a *preliminary* survey module and a compact within-study benchmark, not definitive construct validation or causal evidence on managerial scientific reasoning.

Keywords: scientific decision-making, organizational hierarchy, measurement, evidence-based management, survey instrument

1 Introduction

Innovation research has long used surveys and indicators to make organizational practices visible for research, management, and policy (Gault, 2018; Arundel et al., 2019; Alhusen et al., 2021). Related work measures management and organization (Bloom and Van Reenen, 2010), search for external knowledge (Laursen and Salter, 2004), organizational innovation (Evangelista and Vezzani, 2010), and management practices linked to innovation outcomes (Haneda and Ito, 2018). The premise is not that surveys reveal decision quality directly, but that scoped modules can describe practices at scale.

Scientific decision-making (SDM) creates a related measurement problem. Practices that shape how managers formulate causal claims, evaluate evidence, and revise judgments are plausibly relevant for innovation, but evidence has come mainly from interventions or adjacent literatures rather than from scalable field measurement inside firms (Camuffo et al., 2020, 2024). Firms also differ in how they organize access to knowledge and capture value from it (Cassiman et al., 2010, 2018). This Research Note asks whether a transparent survey module can describe how managers report using evidence-oriented decision practices inside firms.

The natural observational unit is often the working manager in ordinary roles, consistent with evidence that management practices vary within as well as between firms (Bloom et al., 2019). Many upgrading programs operate through firms while targeting individual practices, so whether variation is primarily within or between firms matters for how averages should be read; the module makes that distributional question empirically traceable without equating self-reports with realized decision quality.

This paper introduces a preliminary survey module for reported scientific decision-making practices and applies it to a multi-firm field sample of respondents. The module describes three reported practice domains: Causal Reasoning, Evidence Evaluation, and Adaptive Learning. Diagnostic evidence is intentionally limited in scope and is used to support field deployment rather than to claim definitive construct validation.

As illustrative deployment evidence—rather than as freestanding empirical contributions—

we foreground three patterns from this field application. First, observed SDM scores vary mainly within firms rather than between firms. Second, among hierarchy patterns, Adaptive Learning shows the clearest top-tier advantage, while the full hierarchy pattern is more mixed across dimensions. Third, SDM is positively associated with vignette performance; we treat this as a within-study consistency check, not independent criterion validation (Section 3). The paper does not estimate returns to SDM training, identify promotion mechanisms, or claim effects on firm performance. Its contribution is specific: a field survey module and a compact within-study benchmark for subsequent longitudinal and outcome-linked work. Section 2 defines the construct and scoring architecture; Section 3 summarizes diagnostic evidence; Section 4 reports empirical patterns; Section 5 discusses implications and limits; and Section 6 concludes.

2 Construct and Instrument

Scientific decision-making (SDM) is conceptualized here as a reported organizational practice, enacted by managers and teams, that may help decision-makers recognize, evaluate, and use information under uncertainty in innovation-relevant contexts. This definition connects the "scientific" label to an innovation-studies problem rather than to a generic virtue of rationality: firms benefit from external and internal knowledge only when they can recognize its value, assimilate it, and apply it (Cohen and Levinthal, 1990), and the value of external knowledge depends on complementary internal capabilities (Crescenzi and Gagliardi, 2018). SDM is therefore treated as a manager-level module for observing evidence-oriented practices that may matter for search, experimentation, and learning, not as a complete measure of firm absorptive capacity or realized innovation performance.

The boundary of the construct is equally important. SDM is not generic rationality, general analytical ability, scientific literacy, firm-level absorptive capacity, realized decision quality, or innovation performance. It is a narrower self-report module about whether managers describe their decision practice as involving explicit causal claims, evidence-oriented checks, and willingness to revise commitments. The module can support future work on those broader constructs, but it should not be interpreted as measuring them directly.

Under formative measurement logic, the three domains are not interchangeable reflections of a single latent trait: profiles with high Causal Reasoning but weak Evidence Evaluation, or strong testing without belief updating, are theoretically coherent (Bollen and Lennox, 1991; Jarvis et al., 2003; Diamantopoulos and Siguaw, 2006). It follows that global fit from reflective CFA is not an appropriate primary criterion; poor reflective fit is an expected diagnostic under misspecification rather than a refutation of the item set. Evidence for this module is therefore evaluated through non-redundancy of indicators (variance inflation factors), experimental responsiveness to decision frames, and a within-study vignette consistency check—not through claiming independent reflective validation of three latent factors.

We conceptualize SDM as three analytically separable but potentially recursive practice domains: articulating causal claims, evaluating them against evidence, and updating commitments in light of what is learned. *Causal Reasoning* captures whether decision-makers make explicit claims about how actions are expected to generate outcomes and whether those claims are stated in ways that can in principle be checked against evidence, consistent with cognitive representations that guide attention before full experience is available (Gavetti and Levinthal, 2000); problem discovery in innovation teams motivates extensions but is not a separate headline construct here (Cromwell and Harvey, 2025). *Evidence Evaluation* captures whether managers report designing and interpreting informative tests rather than relying on confirmatory signals, aligned with experimentation as a core innovation activity rather than a generic testing metaphor (Thomke, 1998). *Adaptive Learning* captures reported willingness to revise beliefs, choices, and commitments in response to new evidence, in line with distinctions between generating and evaluating alternatives (Knudsen and Levinthal, 2007), managerial judgment (Bazerman and Moore, 2013), and failures to update when evidence challenges prior commitments (Staw, 1981; March, 1991; Levinthal and March, 1993). The components are conceptually distinct: high Causal Reasoning without Evidence Evaluation yields explanations not disciplined by tests; high Evidence Evaluation without

Adaptive Learning yields testing that does not translate into revised choices. Similar overall scores can mask different profiles, so we interpret dimension scores alongside the composite (Alhusen et al., 2021).

The field instrument has 24 five-point bipolar items (11 Causal Reasoning, 9 Evidence Evaluation, 4 Adaptive Learning). Dimension scores and the overall SDM score are respondent-level means of scored items (reverse-coding and completion rules in Online Appendix Sections A2.1 and A3.4). Qualitative grounding, expert review, the priming-based item screen, six-step design-time mapping, ethics, and full wording appear in Online Appendix Sections A1–A2.

Each item contrasts alternative managerial orientations rather than a single "correct" pole, which reduces transparent acquiescence relative to one-sided agree–disagree stacks, although socially desirable responding remains a concern in employer-sponsored self-reports. The module is designed for organizational deployment rather than laboratory elicitation: it complements behavioral experiments and process tracing by locating variation in *reported* practices at scale across individuals and firms.

3 Diagnostic Evidence

Survey evidence on complex innovation and management practices is vulnerable to wording, interpretation, social desirability, and format (Cirera and Muzi, 2020; Bloom et al., 2016). Under the formative logic in Section 2, the diagnostic questions are narrow: non-redundant indicators, poor reflective fit as an expected misspecification diagnostic, experimental responsiveness to decision frames, and association with a within-study vignette benchmark (MacKenzie et al., 2011; Coltman et al., 2008; Cenfetelli and Bassellier, 2009). Table 1 summarizes results; Online Appendix Sections A4–A5 provide interpretation, CFA detail, priming composition, progressive vignette models, and agreement-style checks.

Each row of Table 1 answers a narrow question and has a clear ceiling. Non-redundant indicators (low VIFs) rule out gross collinearity among formative inputs, but they do not establish

Table 1: Diagnostic Checks and Within-Study Benchmarks

Check	Evidence
Indicator non-redundancy	Item VIFs range from 1.02 to 1.23 across the 24 indicators.
Internal structure	Reflective CFA models fit poorly across one-, two-, and three-factor specifications (three-factor model: CFI = .418, TLI = .355, RMSEA = .054, AIC = 100.37; listwise $N = 2{,}745$).
Experimental responsiveness	Random assignment to scientific, no-prime control, and intuitive decision frames produces strongly ordered SDM scores, with a large overall effect ($F(2, 459) = 175.21$, $p < .001$, $\eta^2 = .43$; $N = 462$).
Vignette consistency	SDM correlates with performance on a short within-study decision vignette ($r = .19$, $p < .001$, $N = 2{,}745$).

Note. VIF = variance inflation factor. SDM = scientific decision-making. CFA = confirmatory factor analysis. Full estimates and caveats appear in the Online Appendix (Sections A4.0–A4.3 and A5).

that the item set is complete or that each dimension label is uniquely correct. Poor fit from reflective CFA specifications is consistent with distinct practice bundles under formative logic, yet it could also reflect misspecified items; it is therefore a diagnostic under an intentionally wrong reflective model, not affirmative evidence of the three-part architecture. The priming exercise documents responsiveness to experimentally induced decision frames, but the scientific frame overlaps the survey content and cannot stand as independent criterion validation (Online Appendix Section A1.2). The vignette row documents association with a differently formatted within-survey task ($r = .19$); in firm-fixed-effects specifications the slope remains positive with full covariates (Online Appendix Section A5.4), while agreement-style checks in Section A5.5 address one response-style channel but not all self-presentation risk. Extended interpretation of these checks, including practice-bundle logic and experimental details, appears in Online Appendix Section A4.0.

In sum, the checks in Table 1 are mutually reinforcing but bounded: none substitutes for external criterion data or longitudinal organizational outcomes. The vignette benchmark is documented in Online Appendix Section A2.2; the association is a within-study consistency check only (same-team instrument, short format; progressive models Section A5.4; agreement-style checks Section A5.5) and does not establish external validity.

4 Empirical Patterns

4.1 Data and Sample

We apply the module to a field sample of 3,178 respondents in 54 firms. We use *respondents* as the generic label. Hierarchy analyses distinguish five tiers from Tier 1 (executives and top management) through Tier 5 (frontline management), based on each respondent's self-placement on the firm's hierarchy question. The survey was a voluntary, web-based organizational field deployment in English, Italian, or Brazilian Portuguese. Participating firms span manufacturing, technology and financial services, pharmaceuticals and biotechnology, and energy and related services, recruited largely through partner organizations with strong ties to Italy and Brazil, although respondents' work locations are not a designed cross-national sample. Firm-level recruitment, HR coordination, invitation metadata, sector composition, translation, and exclusion rules are documented in Online Appendix Section A3.

The empirical exercise is organized around three field-deployment questions. First, does SDM vary mainly between firms or within firms? Second, are any dimensions systematically associated with hierarchical position? Third, do SDM scores correspond to performance on a differently formatted decision task? These questions establish a compact empirical pattern and make a tractable agenda visible without overextending the design.

4.2 Within-Firm Variation

In this sample, under the mixed-effects REML estimator reported in Table 2, most observed variation in reported SDM lies within firms rather than between firms. The estimated ICC for overall SDM is 0.018, implying that 1.8% of variance is attributable to firm membership; subdimension ICCs are 2.2% (Causal Reasoning), 5.4% (Evidence Evaluation), and 1.0% (Adaptive Learning). Overall SDM and Causal Reasoning and Adaptive Learning remain small relative to conventional benchmarks for meaningful group-level clustering (Bliese, 2000), while Evidence Evaluation shows more detectable firm-level clustering in this sample. An equal-weightby-dimension overall score reduces the firm-level ICC (Online Appendix Section A5.2), so the headline conclusion—limited clustering for the overall index—is robust in direction but sensitive in magnitude to aggregation choices. This cross-sectional pattern is compatible with broader evidence that management practices can vary substantially within firms and plants (Bloom et al., 2019). It does not imply that organizations are irrelevant for scientific decision-making; it implies that firm membership alone explains little of the observed score variation in this sample.

Table 2: Between-Firm vs. Within-Firm Variance in SDM (ICC)

	Variance					
Construct	Between-firm	Within-firm	ICC	Between (%)	N	Firms
SDM Overall	0.0015	0.0810	0.018	1.8%	2,745	54
Causal Reasoning	0.0028	0.1243	0.022	2.2%	2,897	54
Evidence Evaluation	0.0097	0.1702	0.054	5.4%	2,988	54
Adaptive Learning	0.0036	0.3491	0.010	1.0%	3,111	54

Note. Random intercept model ($Y_{ij} \sim 1 + (1|\text{firm})$) estimated via REML. ICC(1) = $\sigma^2_{\text{between}}/(\sigma^2_{\text{between}} + \sigma^2_{\text{within}})$. Between (%) is the share of total variance attributable to firm membership. N is the number of managers with non-missing scores for the listed outcome.

4.3 Hierarchy and Adaptive Learning

The clearest hierarchy pattern is not a general seniority gradient. It is a top-tier advantage in Adaptive Learning. This pattern is best interpreted against literatures on where innovation decisions and external-knowledge use sit inside organizations: delegation of innovation decisions depends on innovation strategy and the nature of external knowledge (Colombo et al., 2021), and firms' ability to benefit from external knowledge depends on internal absorptive capacity (Cohen and Levinthal, 1990; Crescenzi and Gagliardi, 2018). Table 3 reports a compact Tier 1-versus-frontline contrast by dimension with and without firm fixed effects; full five-tier estimates and mixed patterns on other tiers and dimensions appear in Online Appendix Sections A5.3 and A5.3b (the latter replicates hierarchy using a permissive overall-score rule). Decision-area controls refer to dummies for corporate strategy, competitive strategy, internal transformation, external stakeholders, and human resources.

Table 3: Tier 1 (Top Management) Versus Frontline: Compact Dimension Summary

Outcome	No Firm FE	Firm FE
Overall SDM	0.04	0.02
	(0.02)	(0.02)
Causal Reasoning	0.05	0.03
	(0.03)	(0.03)
Evidence Evaluation	0.01	-0.03
	(0.03)	(0.03)
Adaptive Learning	0.14***	0.16***
	(0.04)	(0.04)
Controls	Educ., tenure, decision-area	Educ., tenure, decision-area
N	2,745–3,111	2,745–3,111
Firms	54	54

Note. Coefficients shown for Tier 1 (C-suite) relative to Tier 5 (frontline). Estimates are extracted from Panel C of the full five-tier specifications in Online Appendix Table A5.3. No-FE models use robust standard errors; FE models include firm fixed effects and firm-clustered standard errors. Standard errors in parentheses. *** $p < .001$, ** $p < .01$, * $p < .05$.

Tier 1 managers score higher on Adaptive Learning than frontline managers. The unconditional mean difference is $\Delta = .17$ scale points (Tier 1 mean: 3.58; frontline mean: 3.41). With education, tenure, and decision-area controls, the Tier 1–frontline contrast is $b = .14$ ($p < .001$) without firm fixed effects and $b = .16$ ($p < .001$) with firm fixed effects (Online Appendix Table A5.3, Panel C). No comparable top-tier pattern appears for Causal Reasoning or Evidence Evaluation in that specification. The hierarchy pattern is therefore probabilistic, not a clean sorting rule.

4.4 Vignette Performance

SDM is positively associated with performance on the short decision vignette (items in Online Appendix Section A2.2). Respondents in the top SDM quintile select scientifically consistent responses 53.1% of the time, compared with 43.8% for respondents in the bottom quintile (Online Appendix Section A5.4 tabulates quintiles, progressive regressions, and predictive-invariance checks). We treat this solely as a *within-study consistency check*: managers who report more scientific decision practices also perform better on a differently formatted task administered in the same survey. It is *not* independent nomological evidence, criterion validation, or evidence that the module tracks real-world decision quality, because the vignette was developed by the same research team and overlaps conceptually with the SDM construct.

5 Discussion and Implications

This paper proposes a survey module for studying reported scientific decision practices in field settings relevant for research on firms, innovation, and management policy. The empirical contribution is deliberately non-causal and should be understood as part of a broader innovation-measurement agenda in which indicators support monitoring, policy learning, and sharper research questions rather than definitive causal claims (Gault, 2018; Arundel et al., 2019). In this sample, observed SDM scores vary predominantly within firms, Adaptive Learning shows a top-tier advantage relative to frontline managers, and SDM is positively associated with performance on a short decision vignette. These patterns may help motivate future work linking reported decision practices to training interventions and organizational outcomes, but the present design does not test those links (Camuffo et al., 2020, 2024).

For practice, the main implication is caution in interpreting organization-level averages: within-firm variation dominates in this sample, so firm means can mask internal heterogeneity. A dimension-level profile can describe where reported evidence-oriented practices are weaker—especially Adaptive Learning—and provide a baseline for evaluating training or organizational changes aimed at improving belief revision, but it does not support clean ranking of individuals by real decision quality.

The use case is diagnostic and developmental rather than evaluative in a punitive sense. Organizations could administer the instrument before training, compare dimensions across teams or roles, and use the pattern to decide whether to emphasize hypothesis formulation, evidence design, or belief updating. The evidence does not support treating SDM scores as a personnel instrument; they are better read as a starting point for reflection and intervention design.

For policy and intermediary programs, the module may eventually serve as a baseline for evidence-based management support, contingent on stronger independent validation. Reasonable uses include describing baseline practices, stratifying interventions by dimension, and tracking whether programs change the practices they target. That bounded use is especially salient for programs that combine managerial upgrading with innovation or technology-adoption support, where implementers must judge whether constraints lie in lack of information, weak causal diagnosis, poor test design, or resistance to updating after evidence arrives. A dimension-level profile cannot answer that diagnosis alone, but it can offer a common baseline for tailoring support and for comparing pre- and post-intervention changes in reported practices. Used this way, the module complements richer process data, interviews, or outcome tracking. Visibility at scale is not causal attribution: the module fits program design where the target is a change in practices; productivity or innovation-outcome claims require independent data (Aghion and Tirole, 1997; Joseph and Gaba, 2020; Cassiman et al., 2010, 2018; Crescenzi and Gagliardi, 2018).

For researchers, the top-tier Adaptive Learning advantage is the main empirical puzzle. It is not a general hierarchy gradient across all SDM dimensions, and it persists after conditioning on education and tenure. Because those controls absorb much of the human-capital credentialing channel, a story in which formal schooling or tenure alone produces the Tier 1 contrast on Adaptive Learning is at best incomplete: the residual hierarchy pattern is not plausibly driven *solely* by credentials that are already in the model. Firm-wide cultural homogenization is also an unparsimonious explanation for this specific finding: Adaptive Learning shows negligible firm-level clustering in this sample (intraclass correlation about 0.010), so most of the cross-manager variation—and therefore the hierarchy gradient—is within firms rather than reflecting a single shared organizational culture that lifts all tiers equally. What remains most consistent with both the low firm-level clustering and the dimension specificity (Adaptive Learning, not Causal Reasoning or Evidence Evaluation) is *role-based exposure*: senior roles disproportionately involve novel, high-ambiguity problems where beliefs must be revised as evidence accumulates, whereas frontline roles more often execute routines with less visible scope for updating. Selection into senior roles and post-promotion socialization can still be confounded with exposure in cross-section. Longitudinal designs that follow the same managers through promotions or lateral moves—ideally paired with administrative job descriptions or project portfolios—are needed to test whether Adaptive Learning rises when job content becomes more ambiguous (exposure) rather than whether managers who already update readily are sorted upward (selection). The delegation literature nonetheless emphasizes that where innovation decisions and external knowledge sit in hierarchies varies with strategy and context (Colombo et al., 2021), so the pattern should be read as a prompt for theory on role design rather than as a settled attribution.

The evidence has important limits. The field sample is voluntary and mostly Italian, limiting generalizability. Because how hierarchy maps onto decision rights, delegation of novel problems, and accountability for belief revision varies across national institutional contexts, the Adaptive Learning hierarchy gradient is plausibly culturally and legally sensitive (Colombo et al., 2021). Until the module is replicated in more diverse country portfolios, that finding should be treated as an Italian- and European-context regularity rather than as a universal feature of managerial cognition. Like other survey modules on complex management and innovation practices, this instrument trades depth for standardized coverage and remains vulnerable to respondent interpretation, framing, and self-presentation (Bloom et al., 2016; Cirera and Muzi, 2020). The priming experiment is a sensitivity check rather than stand-alone construct validation; longitudinal evidence and independently designed criteria remain necessary. The vignette benchmark is short and was developed by the research team. The SDM–vignette association is therefore a useful first benchmark, not evidence that the module predicts real organizational outcomes. Independent external validation would require linking the survey to observed decisions with verifiable ex-post outcomes (for example project go/no-go choices tied to documented performance), to multi-rater assessments of decision quality by role (peer or supervisor ratings in designs analogous to 360 feedback), and to project-level innovation indicators from administrative data. Future work should pursue those links alongside independently designed vignettes and longitudinal organizational outcomes.

6 Conclusion

This note contributes a 24-item field module and a compact within-study benchmark for reported scientific decision-making practices, not a validated performance metric or causal account of hierarchy. Under the present scoring rule, the field deployment illustrates three patterns: most variation in SDM lies within firms, Adaptive Learning shows a top-tier advantage relative to frontline managers, and SDM covaries with a short within-survey vignette benchmark.

The contribution is a measurement tool for describing reported practices and where they vary, not a definitive account of managerial scientific reasoning. In that sense the module sits closer to an innovation-measurement agenda for policy learning and future research than to a performance index (Gault, 2018; Arundel et al., 2019). It does not resolve why Adaptive Learning is more common among top-tier managers or whether the module predicts consequential organizational outcomes; longitudinal designs, independent criteria, and administrative outcomes remain the next empirical steps.

References

Aghion, P., Tirole, J., 1997. Formal and real authority in organizations. Journal of Political Economy 105, 1–29. doi:10.1086/262063.

Alhusen, H., Bennat, T., Bizer, K., Cantner, U., Horstmann, E., Kalthaus, M., Proeger, T., Sternberg, R., Töpfer, S., 2021. A new measurement conception for the 'Doing-Using-Interacting' mode of innovation. Research Policy 50, 104214. doi:10.1016/j.respol.2021.104214.

Arundel, A., Bloch, C., Ferguson, B., 2019. Advancing innovation in the public sector: Aligning innovation measurement with policy goals. Research Policy 48, 789–798. doi:10.1016/j.respol.2018.12.001.

Bazerman, M.H., Moore, D.A., 2013. Judgment in Managerial Decision Making. 8 ed., John Wiley & Sons, Hoboken, NJ.

Bliese, P.D., 2000. Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis, in: Klein, K.J., Kozlowski, S.W.J. (Eds.), Multilevel Theory, Research, and Methods in Organizations. Jossey-Bass, pp. 349–381.

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I., Van Reenen, J., 2019. What drives differences in management practices? American Economic Review 109, 1648–1683. doi:10.1257/aer.20170491.

Bloom, N., Lemos, R., Sadun, R., Scur, D., Van Reenen, J., 2016. International data on measuring management practices. American Economic Review 106, 152–156. doi:10.1257/aer.p20161058.

Bloom, N., Van Reenen, J., 2010. New approaches to measuring management and firm organization. American Economic Review 100, 105–108.

Bollen, K.A., Lennox, R., 1991. Conventional wisdom on measurement: A structural equation perspective. Psychological Bulletin 110, 305–314. doi:10.1037/0033-2909.110.2.305.

Camuffo, A., Cordova, A., Gambardella, A., Spina, C., 2020. A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. Management Science 66, 564–586. doi:10.1287/mnsc.2018.3249.

Camuffo, A., Gambardella, A., Messinese, D., Novelli, E., Paolucci, E., Spina, C., 2024. A scientific approach to entrepreneurial decision-making: Large-scale replication and extension. Strategic Management Journal 45, 1209–1237.

Cassiman, B., Di Guardo, M.C., Valentini, G., 2010. Organizing links with science: Cooperate or contract? Research Policy 39, 882–892. doi:10.1016/j.respol.2010.04.009.

Cassiman, B., Veugelers, R., Arts, S., 2018. Mind the gap: Capturing value from basic research through combining mobile inventors and partnerships. Research Policy 47, 1811–1824. doi:10.1016/j.respol.2018.06.015.

Cenfetelli, R.T., Bassellier, G., 2009. Interpretation of formative measurement in information systems research. MIS Quarterly 33, 689–707. doi:10.2307/20650323.

Cirera, X., Muzi, S., 2020. Measuring innovation using firm-level surveys: Evidence from developing countries. Research Policy 49, 103912. doi:10.1016/j.respol.2019.103912.

Cohen, W.M., Levinthal, D.A., 1990. Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly 35, 128–152. doi:10.2307/2393553.

Colombo, M.G., Foss, N.J., Lyngsie, J., Rossi Lamastra, C., 2021. What drives the delegation of innovation decisions? the roles of firm innovation strategy and the nature of external knowledge. Research Policy 50, 104134. doi:10.1016/j.respol.2020.104134.

Coltman, T., Devinney, T.M., Midgley, D.F., Venaik, S., 2008. Formative versus reflective measurement models: Two applications of formative measurement. Journal of Business Research 61, 1250–1262. doi:10.1016/j.jbusres.2008.01.013.

Crescenzi, R., Gagliardi, L., 2018. The innovative performance of firms in heterogeneous environments: The interplay between external knowledge and internal absorptive capacities. Research Policy 47, 782–795. doi:10.1016/j.respol.2018.02.006.

Cromwell, J.R., Harvey, J.F., 2025. A problem half-solved is a problem well-stated: Increasing the rate of innovation through team problem discovery. Research Policy 54, 105186. doi:10.1016/j.respol.2025.105186.

D'Este, P., Iammarino, S., Savona, M., von Tunzelmann, N., 2012. What hampers innovation? revealed barriers versus deterring barriers. Research Policy 41, 482–488. doi:10.1016/j.respol.2011.09.008.

Diamantopoulos, A., Siguaw, J.A., 2006. Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. British Journal of Management 17, 263–282. doi:10.1111/j.1467-8551.2006.00500.x.

Evangelista, R., Vezzani, A., 2010. The economic impact of technological and organizational innovations: A firm-level analysis. Research Policy 39, 1253–1263. doi:10.1016/j.respol.2010.08.004.

Gault, F., 2018. Defining and measuring innovation in all sectors of the economy. Research Policy 47, 617–622. doi:10.1016/j.respol.2018.01.007.

Gavetti, G., Levinthal, D., 2000. Looking forward and looking backward: Cognitive and experiential search. Administrative Science Quarterly 45, 113–137. doi:10.2307/2666981.

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

Haneda, S., Ito, K., 2018. Organizational and human resource management and innovation: Which management practices are linked to product and/or process innovation? Research Policy 47, 194–208. doi:10.1016/j.respol.2017.10.008.

Jarvis, C.B., MacKenzie, S.B., Podsakoff, P.M., 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. Journal of Consumer Research 30, 199–218. doi:10.1086/376806.

Joseph, J., Gaba, V., 2020. Organizational structure, information processing, and decision-making: A retrospective and road map for research. Academy of Management Annals 14, 267–302. doi:10.5465/annals.2017.0103.

Knudsen, T., Levinthal, D.A., 2007. Two faces of search: Alternative generation and alternative evaluation. Organization Science 18, 39–54. doi:10.1287/orsc.1060.0216.

Laursen, K., Salter, A., 2004. Searching high and low: What types of firms use universities as a source of innovation? Research Policy 33, 1201–1215. doi:10.1016/j.respol.2004.07.004.

Levinthal, D.A., March, J.G., 1993. The myopia of learning. Strategic Management Journal 14, 95–112. doi:10.1002/smj.4250141009.

MacKenzie, S.B., Podsakoff, P.M., Podsakoff, N.P., 2011. Construct measurement and validation procedures in mis and behavioral research: Integrating new and existing techniques. MIS quarterly 35, 293–334.

March, J.G., 1991. Exploration and exploitation in organizational learning. Organization Science 2, 71–87. doi:10.1287/orsc.2.1.71.

Staw, B.M., 1981. The escalation of commitment to a course of action. Academy of Management Review 6, 577–587. doi:10.5465/amr.1981.4285694.

Thomke, S.H., 1998. Managing experimentation in the design of new products. Management Science 44, 743–762. doi:10.1287/mnsc.44.6.743.