Supply Chain Research · Cornerstone Reference

Scoring Implementation Risk: A Method

How Do I Assess the Risk of a Supply Chain Software Implementation?

Score six dimensions before approval, not one. The useful output is the shape of the exposure rather than a single index, because a project that is average overall and severe on one dimension is the one that fails.

CONTENTS

01 The short answer

Assess implementation risk by scoring six dimensions before approval: scope stability, integration complexity, data readiness, change load, vendor viability, and internal team capacity. Score each on a coarse scale with written evidence, and read the profile rather than the total. A project scoring moderately across all six is usually manageable, while a project that averages well but scores severely on one dimension is the one that fails, because a single unmitigated exposure is enough to stop a go-live.

6	1 to 5	1
dimensions, scored separately and never collapsed into one index	a deliberately coarse scale, because finer resolution is false	named owner and mitigation required per dimension above threshold

Key takeaways

- **Read the shape, not the total.** Averaging hides the single severe exposure that actually stops projects.
- **This is project risk, not supply chain risk.** It concerns whether the implementation succeeds, not whether suppliers or networks fail.
- **Score with evidence or do not score.** Each rating should cite something observable, because unsupported scores drift toward comfortable answers.
- **Keep the scale coarse.** A five-point scale is honest about the precision available; a hundred-point index is not.
- **Scores should trigger action, not commentary.** Any dimension above threshold needs a named owner and a mitigation before approval.

02 Which dimensions should be scored?

Six cover most of what goes wrong. Scope stability asks whether the boundary is written, agreed, and unlikely to move, since scope that is still being negotiated at kickoff will keep moving. Integration complexity counts the interfaces to systems that are not being replaced, weighted by whether each is a supported connector or bespoke work. Data readiness asks whether the master data the new system requires exists at the quality it demands, which is usually discovered late and is rarely negotiable.

Change load asks how many people must work differently, across how many sites, and whether the operating model itself changes or only the tooling. Vendor viability covers financial stability, ownership changes, product roadmap for the specific module being bought rather than the suite, and whether the version being sold is the one under active development. Internal team capacity asks whether the named people are actually available, which in practice means whether they are also running the operation or delivering other projects during the same period.

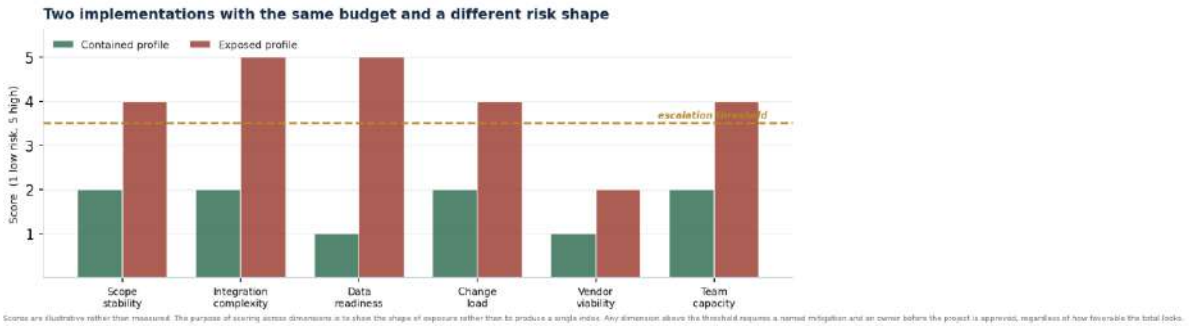

Figure 1. Two illustrative projects with comparable budgets and different exposure shapes. The contained profile is manageable. The exposed profile has three dimensions above threshold, and the total conceals that.

03 How do I score them without inventing precision?

Use a five-point scale and require written evidence for each rating. Coarse scales are more honest than fine ones here, because the information available at approval does not support finer resolution, and a weighted composite calculated to one decimal place implies a confidence nobody has. The evidence requirement matters more than the scale: a score of two on data readiness should cite a profiling result or a known defect rate, not an impression.

Score with the people who will do the work rather than only with the sponsor, and record disagreement rather than resolving it into an average. Where an implementation lead scores integration at four and the sponsor scores it at two, the gap is the finding. Averaging it to three destroys the only useful information in the exercise, which is that the two people accountable for the outcome hold incompatible views of the same project.

The fair objection is that scoring of this kind is subjective and can be shaped to produce a desired conclusion. That is true of any pre-project assessment, and the mitigation is procedural rather than mathematical: require evidence, record dissent, and have someone outside the project team review the ratings. None of that makes the scores objective. It makes them auditable, which is the achievable standard.

Dimension	What raises the score	Evidence to demand
Scope stability	Boundary still under negotiation at kickoff	A signed scope document and a written change process
Integration	Many interfaces, bespoke rather than supported connectors	An interface inventory naming each system, owner, and method
Data readiness	Master data below the quality the new system requires	A profiling result against the target system's field requirements
Change load	Many users, many sites, operating model changing not just tooling	A count of affected roles by site and a named change owner
Vendor viability	Ownership change, thin roadmap for the specific module	Roadmap for the module bought, and the version under active development
Team capacity	Named people also running operations or other projects	Named individuals with committed allocation, and their backfill plan

Table 1. The six dimensions, what drives each score upward, and the evidence a rating should cite. A score without the third column is an opinion.

04 What do the scores actually trigger?

Set a threshold before scoring, and attach a consequence to crossing it. On a five-point scale, four or above should require a named owner, a written mitigation, and a date by which the score is expected to fall. Two or more dimensions above threshold should trigger a scope or sequencing conversation rather than a mitigation plan, because that pattern usually indicates the project is too large for a single phase.

The consequence needs to be real. An assessment that produces a risk register nobody revisits is worse than no assessment, because it creates the appearance of diligence and consumes the attention that would otherwise go to the exposure. The test of whether the exercise is working is whether a score has ever changed a decision: delayed an approval, split a phase, added a resource, or removed a site from a first wave.

05 How is this different from supply chain risk management?

Different subject entirely, and the two are frequently confused because they share a word. Supply chain risk management concerns exposure in the operation: supplier failure, disruption, concentration, geopolitical and logistics risk. It is an ongoing discipline supported by its own software category. Implementation risk concerns whether a project to change systems succeeds, and it is assessed at approval and reassessed at phase boundaries.

The distinction has a practical consequence in governance. Implementation risk belongs to the project sponsor and the steering group, and it expires when the project closes. Supply chain risk belongs to the operating organization permanently. Placing implementation risk on the operational risk register tends to leave it there long after it stopped being actionable, while the reverse error leaves genuine operational exposure with a project team that disbands.

06 When should the assessment be repeated?

At approval, at each phase boundary, and after any event that changes the answer to one of the six questions. Scope changes, a vendor acquisition, the loss of a named team member, or a data profiling result that arrives later than planned each warrant a rescore of at least the affected dimension. The value comes from the comparison across time rather than from any single assessment, since a dimension moving in the wrong direction across two reviews is a clearer signal than any absolute score.

Keep the rescore short. An assessment that takes a week will not be repeated, and the discipline depends on repetition. Once the first full scoring is complete with its evidence, subsequent reviews should be a working session that confirms what has moved and why.

07 Frequently asked questions

Should we weight the dimensions?

Generally no. Weighting invites the composite score that this method deliberately avoids, and the dimension that matters most varies by project rather than being knowable in advance. If weighting is required by a governance standard, publish the weights before scoring and still report the individual dimensions alongside the total.

What score means we should not proceed?

No threshold answers that on its own, and treating one as decisive would be false precision. The pattern worth stopping for is several dimensions above threshold with no credible owner for the mitigations, which indicates the organization lacks the capacity to absorb the project rather than that the project is badly chosen.

Who should run the assessment?

Someone with no delivery stake in the outcome, working with the people who will deliver. A project manager scoring their own project will produce scores that are defensible rather than accurate, which is a predictable consequence of the incentive rather than a failure of integrity.

How does this relate to the vendor's implementation methodology?

It sits before it. Vendor methodologies describe how a project runs once approved and generally assume the project is viable. This assessment tests whether it is, and several of its dimensions, particularly team capacity and data readiness, are outside the vendor's control and therefore outside their method.

Is a phased rollout always lower risk?

Not automatically. Phasing reduces change load per phase and creates learning between them, but it extends the period of parallel running, which carries its own cost and integration burden. The honest comparison is between a shorter high-load project and a longer one with sustained dual operation.

Should the vendor see the scores?

Selectively. Sharing integration and data readiness scores usually improves planning, because the vendor can resource against a known problem. Sharing team capacity and internal change scores is a judgment about the commercial relationship, since those dimensions describe your organization rather than the work.

What if leadership overrules a high score?

That is a legitimate decision and the method still did its job. The purpose is to make the exposure visible and owned, not to hold a veto. What matters is that the override is recorded with a rationale, because the same assessment is what allows the organization to learn from the outcome either way.

08 Method, sources, and where to go deeper

Method

- This page sets out a scoring method rather than a survey of implementation outcomes. The six dimensions and the threshold rules reflect SCR's own framing of practitioner method.
- The separation between implementation risk and operational supply chain risk follows the process boundaries in the SCOR Digital Standard, which places risk management within Orchestrate as a continuing governance activity rather than a project activity.
- Supply Chain Research is independent and vendor-neutral. We accept no payment from the vendors or categories covered, and this page names no products.

Caveats

- Figure 1 uses illustrative scores to demonstrate profile shape. They are not measured data and should not be read as typical values for any category of project.
- No reliable public dataset exists on supply chain implementation failure rates. Widely quoted failure percentages for enterprise software circulate without consistent definitions of failure, without sample descriptions, and frequently from parties selling remediation services. SCR does not reproduce them.
- Scoring of this kind is judgment-based and cannot be made objective. The procedural controls described in section 03 make it auditable, which is a different and more achievable claim.

Where to go deeper

Two SCR guides carry the category detail this assessment assumes. The warehouse management guide covers the integration and data requirements typical of execution implementations, which is where the integration and data readiness dimensions usually bite hardest. The supply chain planning guide covers the change load characteristic of planning deployments, where the operating model shift is often larger than the technical one. Readers assessing a project that has not yet selected a product should read the SCR selection framework first.

Sources

[1] Association for Supply Chain Management. SCOR Digital Standard overview.

[2] Association for Supply Chain Management. SCOR DS model reference, Orchestrate process including risk mitigation.

[3] Association for Supply Chain Management. Introduction and front matter, SCOR Digital Standard, 2025 edition.