From VP and SE to Decisions, Follow-up, and Learning

Training materials 3.1 | Calculation model 3.0 | US edition: October 7, 2026

Original methodology published September 17, 2026. https://enableaisi.com/insights/did-we-realize-the-expected-value/

Instructor article · Editorial edition 3.1 · Calculation model 3.0 · EnableAI · September 17, 2026

1. The goal is a better decision

VP describes an expected improvement using shared reference points. SE describes all remaining effort using a fixed resource unit. **Priority indicator = expected VP / remaining SE.** Value and size must refer to the same scope and agreed measurement period. This EnableAI formula supports decisions; it is not a financial benefit-cost ratio or a guarantee of the best portfolio.

Before prioritizing, management must agree on the decision at hand: investigate an idea, run a limited test, expand a proven solution, or use the resources elsewhere. Always show both the cost of the next phase and the expected resources needed for the entire remaining path to the promised result. Otherwise, a low-cost pilot can receive an artificially high priority.

Start with the VP article and the SE article. This article connects the calculations with decision meetings, accountability, and follow-up. The interview and calculation rules are EnableAI's practical implementation. The broader principle of connecting pre-investment assessment with subsequent evaluation draws on sources including HM Treasury, Green Book 2026.

2. What belongs in management's project overview?

Show the project, current decision phase, documented scope, central VP estimate, direct and indirect subtotals, low and high scenarios, remaining SE split into K/R/U, VP/SE, next-phase resources, cash requirements, scarce skills, time to benefit, mandatory requirements, and dependencies. Include the accountable owner, limitations of the evidence, most recent review date, and proposed next action.

Flag missing material inputs. An automatic ranking must not conceal an incomplete evidence base. Record the ranking with a brief rationale: “Investigate P4 first because its central ratio is high, business capacity is available, and a short test can clarify the churn effect.” That is easier to evaluate than “P4 scored the most points.”

The five recurring examples before hypothesis testing

Project	VP	SE	VP/SE	Rank based on central estimate
Redesigned customer onboarding	17.00	8.34	2.04	1
Customer service knowledge assistant	19.50	11.47	1.70	2
Invoice processing agent	25.32	18.99	1.33	3
Better IT operations and incident readiness	18.72	18.56	1.01	4

Project	VP	SE	VP/SE	Rank based on central estimate
Shared, secure AI knowledge platform	21.50	28.48	0.75	5

The figures are illustrative model calculations from the five complete project examples. Onboarding has the highest ratio, but the invoice agent has more total VP. The platform may be a prerequisite for other initiatives, so it also needs assessment as part of a combined package. A difference in decimal places is not a reliable distinction when the ranges overlap.

3. A fixed local scale, not a universal truth

Adding a project with 15 VP and 5 SE gives it 3 VP/SE. Existing projects retain their VP and SE, but their ranking may change. This requires consistent criteria, weights, reference points, price year, and measurement period. Normalizing scores against “the best project on today's list” would break this property, so the model does not use that approach.

The scale is more stable than relative project points, but it remains a local preference index. Management chooses how much to value the improvements and how to combine them linearly. Other reasonable preferences can materially change the ratio. Show sensitivity, especially when projects have very different shares of indirect value.

The highest ratio does not always make the best use of a limited budget. With 10 SE available, a project worth 12 VP for 4 SE has a higher ratio than one worth 25 VP for 10 SE. If the remaining 6 SE cannot be used elsewhere, the larger project delivers more total VP. Dependencies, scarce skills, mandatory requirements, and time sensitivity must therefore inform the final portfolio decision.

SAFe's WSJF offers useful ideas for considering flow, time sensitivity, risk reduction, and opportunities. VP/SE does not automatically apply a cost-of-delay formula. Record specific deadlines and the consequences of waiting alongside the ratio. Avoid arbitrary “urgency points” that silently change the model. SAFe, WSJF.

4. Documentation before each phase

Before Hypothesis: Describe the problem, the alternative without the project, an initial benefit map, and an estimate for the entire remaining project. Agree on the uncertain assumption the investigation must resolve. The phase budget funds investigation, not a claim to the full benefit. For example, test whether relevant volume, net time savings, or integration feasibility can be substantiated.

Before MVP: Use hypothesis findings to update VP and SE estimates. Define actual users, cases, quality requirements, and a representative trial. Show both the local effect in the trial and the forecast for full adoption. A small trial can substantiate time per task; it does not automatically substantiate total volume, adoption, or organization-wide readiness.

Before Scaling: Use the MVP's observed results, actual resource use, and organizational lessons. Examine what changes with new user groups, additional integrations, support, permissions, training, and operations. Scaling involves more than adding licenses. Reassess full VP, remaining K/R/U, and the next funding request.

After each phase: Save observations, actual resource use, and explanations for variances. Decide whether to stop, adjust, investigate further, or continue. If the final measurement period cannot yet be observed, label benefits “forecast” or “pending.” Do not relabel a proxy as realized revenue or savings.

5. What should a manager ask at a phase review?

Question	Required evidence	Possible action
What have we actually learned?	New observations with sources and scope	Update specific inputs, not just the total
Which benefits have been measured, and which are still expected?	Underlying indicators and measurement period	Separate actual results from forecasts
How accurate were our estimates?	Frozen target, tolerance, and consistent measurement method	Learn from errors without rewriting history
Is the result good enough?	Minimum benefit and mandatory requirements	Do not continue simply because the estimate was accurate
What will it take from here?	Remaining activities, purchases, R, and U	Make a new funding and capacity decision
What happens if we stop or choose another option?	Alternatives, commitments, and any shutdown work	Decide based on future differences
Who acts next, and when?	Owner, date, conditions, and review	Turn the decision into a concrete action

The accountable business owner accepts the benefit target; the project manager owns consistency and version history; subject-matter leads stand behind their estimates; and finance reviews the pricing assumptions and budget effects. A shared portfolio function brings together lessons across projects and tracks benefits after delivery. This can be a few people meeting regularly, rather than a new organizational department.

6. Freeze the baseline and record changes explicitly

Save each approved scorecard and phase estimate with a decision ID, date, scope, model profile, formulas, data sources, and approvers. New evidence creates a new forecast; retain the old one. The spreadsheet's Phase Follow-up and Measurement Points tabs contain frozen inputs. New figures in the calculation tabs do not automatically replace them.

Log every change with the old figure, new figure, reason, evidence, and decision. For example: “Usage share reduced from 80% to 65% because two groups will not have access in the first 90 days; supported by the access plan; affects D1 and training scope.” A change can reduce VP and also change SE. Do not adjust all figures using a blanket maturity allowance.

When calibration changes, retain both the original and new profiles. Measure historical accuracy against the originally approved target. Recalculate current priorities using one shared profile. Do not combine old VP from a previous weighting with new SE without identifying and reconciling the difference.

7. How to calculate variances

Value estimation error: (actual improvement - expected improvement) / absolute expected improvement. If time was expected to fall from 12 to 9 minutes, the expected improvement is 3. An actual reduction to 9.6 minutes is an improvement of 2.4, giving a variance of -20%. Calculating the error against the final level of 9 minutes would be incorrect.

Size estimation error: (scope-adjusted actual resource use - approved phase estimate) / approved phase estimate. Show the unadjusted difference alongside it. A scope adjustment requires a separately documented change. Forgotten work or an anticipated risk that occurred does not justify rewriting the baseline.

Final forecast error: Compare final resource use with both the original total and subsequent total forecasts. Also show the variance from the base estimate, so a large reserve does not make a poor activity estimate appear accurate. An

unused reserve is not a delivered benefit in VP.

Tolerance: Accepted absolute difference = the greater of |target| × relative tolerance and an agreed absolute tolerance. For a zero target, use the absolute difference; do not divide by zero. The examples use ±50/25/15% as illustrative phase tolerances. These are management rules, not statistical confidence intervals or industry benchmarks.

Minimum benefit and estimation tolerance answer different questions. If the target is 3 minutes, tolerance is ±25%, and the minimum is 2.5, a result of 2.4 minutes is within tolerance but below the minimum. The decision must address the inadequate benefit. Higher VP elsewhere cannot compensate for failure to meet a safety requirement either.

8. Learn from accurate and inaccurate estimates

Break down the explanation into volume, benefit per unit, adoption, use of released capacity, price, productivity, scope, and risk events. Look for repeated errors. If projects consistently underestimate participation time, update the question guide and reference data. Do not recalibrate the model after a single unusual event without checking comparability.

Review several completed projects using consistent definitions. Show both the direction and size of variances, and include projects that were stopped. Otherwise, you risk learning only from successes. Document any new reference profile with a date and rationale. GAO supports updating cost estimates with actual resource use; this specific learning cycle is EnableAI's recommendation. GAO, 2020.

Also examine whether tests actually reduce uncertainty in decisions. A hypothesis investigation can succeed even when it leads management to stop: it has prevented a larger investment based on a false assumption. Describe that learning specifically. Do not automatically award extra VP to compensate for a poor project.

9. Teaching interpretation: three decision exercises

Exercise 1: A new project. Add a project with 15 VP and 5 SE to the comparison. Answer: a ratio of 3.00 and the highest rank under the central estimates. Existing VP/SE scores do not change. Then ask participants to identify a capacity constraint that could still prevent the project from starting.

Exercise 2: Overlapping scenarios. Project A has central estimates of 20 VP and 10 SE; B has 16 VP and 9 SE. In a coherent unfavorable scenario, A has 10 VP and 15 SE, while B has 14 VP and 10 SE. Answer: A ranks first under the central estimates, 2.00 versus 1.78; B ranks first in the unfavorable scenario, 1.40 versus 0.67. Identify the assumption that is both most important and least expensive to clarify before choosing.

Exercise 3: Past resource use. A project has already used 8 SE. It requires another 4 SE to create 10 future VP that would not result from stopping. An alternative requires 4 SE and produces 12 future VP. Answer: past spending does not make the first project more attractive. Compare future alternatives, dependencies, shutdown costs, and risk. Retain the original project ratio for follow-up.

10. Sources, examples, and skills assessment

The cited sources support the principles, not EnableAI's numeric calibration. The Morgan Stanley case can support a discussion of ongoing quality control. METR's study can support a discussion of perceived versus measured results. The five illustrative project examples show concrete calculations without presenting them as documented customer results.

An instructor should be able to run a phase review, reproduce VP and SE from source evidence, spot two types of double counting, distinguish tolerance from minimum benefit, and explain why the highest-ranked option may not be selected. Use the workshop guide and the templates to practice these skills.