Business Value Points in Practice

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

Original methodology published September 17, 2026. https://enableaisi.com/insights/how-to-measure-business-value/

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

Reading guide and learning objectives

This article is a working guide for project managers, product owners, business owners, and instructors. Start with sections 1-4, work through the criteria in sections 5-6, then finish with calculation, measurement, and training exercises. Use the size article to estimate effort, and the decision and follow-up article for overall prioritization. The two half-day workshops turn the material into shared practice.

After working through the material, you should be able to interview an accountable business owner, distinguish outcomes from activities, document a baseline, calculate net improvements, and explain exactly why a project receives its VP. You should also be able to reject unsupported figures, identify double counting, and teach others to reassess an estimate. Reading alone does not make an experienced instructor. Complete the exercises, seek expert feedback, and try the method with your organization's own data.

This guide deliberately goes into more detail than the main website. Its numerical examples are training exercises. Publicly documented customer cases are identified separately; they do not substantiate EnableAI results or the model's accuracy.

1. What comes from the sources, and what has EnableAI chosen?

Source-based principle. Multi-Criteria Decision Analysis (MCDA) provides a basis for comparing different outcomes through explicit criteria, preferences, and sensitivity analysis. Weights must relate to the size of the improvement being compared, not simply the name of a criterion. S1: Government Analysis Function, MCDA.

EnableAI's practical implementation. The 12 criteria, their grouping, reference improvements, starting weights, 90-day window, the name VP, and the linear conversion below are our methodological choices. They are not a certified MCDA standard or empirically validated exchange rates. The organization must accept them as a shared preference profile and test whether the ranking is robust enough for the decision.

Source-based inspiration for indirect value. SAFe's WSJF includes risk reduction and new opportunities in cost of delay. EnableAI draws on this idea to make future effects visible. VP/SE is not WSJF: its denominator is normalized resource use, and time sensitivity is not automatically included in the numerator. S2: SAFe, WSJF.

EnableAI's teaching approach. The interview sequence, scorecards, evidence requirements, examples, and competency assessment in this article form our recommended workflow. They support a Business Agility principle: fund bounded steps, get feedback from real workflows, and change the decision when the evidence changes. The specific phases are EnableAI's choice.

2. Start with a decision and a comparison baseline

Do not start by asking, “How valuable is this idea on a scale of 1 to 10?” Ask, “Who should be able to do what better, compared with which realistic alternative, by when, and under what conditions?” Without a shared reference point, the same score means different things to different participants.

Agree on the project's scope, affected tasks, included units, and expected adoption date. The baseline is the observed starting point. The no-project alternative is the expected course without the proposed investment; it may differ from the baseline if, for example, volume is already rising or another improvement has been approved. Only the difference relative to that alternative should be attributed to the project.

By default, use the first 90 days after the agreed adoption date. Specify exact dates, expected rollout, and volume during that period. Do not imply that full future capacity will be available on day one. A 90-day window is a comparison convention, not evidence that long-term investments are inferior. Projects with substantially different lifetimes, delays, or ongoing operating burdens also require a consistent lifecycle analysis.

Here, direct value means an observable improvement from using the deliverable. Indirect value means a documented foundation for later benefits. The distinction concerns the nature of the benefit, not whether business or IT commissioned the project. A reliable integration can provide direct operating value to customer service. A reusable data model can create indirect value for future projects.

The interview: eight questions that must be answered

Question	How to document the answer	Effect on the calculation
Which specific workflow changes?	Process step, start and end, task type, and exceptions	Defines which tasks may be counted
Who benefits from the improvement?	Beneficiary group and named value owner	Defines the population to measure and who is accountable
What happens without the initiative?	Baseline period, data source, and expected trend	Provides a relevant comparison
What becomes measurably better?	Natural unit, direction, and definition	Determines the criterion and positive or negative sign
How often does it happen?	Volume during the measurement period and seasonal variation	Converts benefit per task into total benefit
How much will actually be used?	Usage share, rollout, and use of released time	Defines the improvement that can actually be realized
What could reduce the benefit?	Review, rework, other adverse effects, and quality requirements	Net calculation and scenarios
How will we know if we were wrong?	Measurement method, timing, minimum, and tolerance	Makes the estimate testable

An interview can be prepared in 30-45 minutes, but that is not a deadline for obtaining missing data. It is better to finish with three investigations and named owners than with three guessed scores.

3. A documented answer is more than a number

Create one scorecard per independent benefit, such as V-P1-D1-01. Include the project and phase IDs, model version, criterion, accountable owner, beneficiary, measurement period, baseline, no-project alternative, low/central/high expected improvement, unit, complete formula, data sources, measurement method, limitations, and next review. Also record the minimum benefit, tolerances, and links to potentially overlapping benefits.

For each input, record whether it is **observed**, **calculated**, **drawn from a comparable project** or **an expert assumption**. These are EnableAI's evidence categories, not a mathematical confidence score. An observed figure may still contain errors or be unrepresentative. An assumption can be used in an explicit scenario, but it needs an owner responsible for testing it.

Do not write only “CRM shows 12,000 cases.” Write, for example: “CRM-Q3 extract, version 2, queues A-C, creation date, duplicates removed by case ID, closed test queue excluded; 12,000 relevant cases expected in the selected 90-day window.” Save the extract or query, definition, and access to the source. Link the spreadsheet to this evidence; sensitive personal information does not need to be copied into the calculation sheet.

Zero means “assessed as no effect.” Blank means “not investigated.” Explain why something is not applicable before assigning zero. Record adverse effects as negative values. If a material benefit is unknown, the total is incomplete; do not present it as a confirmed zero or a final prioritization.

4. How conversion to VP works

VP for criterion j = weight j × net improvement j / reference improvement j. Total VP is the sum of independent contributions from the criteria. Show direct and indirect subtotals separately. Calculate with full precision and normally display no more than two decimal places.

ID and criterion	Reference improvement	Starting weight	VP at the reference improvement
D1 Capacity put to use	600 net hours	15	15
D2 Existing expense avoided	$22,600	15	15
D3 Quality	60 material errors	10	10
D4 Operational continuity	3,000 lost user hours	10	10
D5 Customer acquisition and retention	30 net customers	10	10
D6 Resource and climate impact	10 metric tons CO2e	5	5
I1 Reusable learning	4 validated instances of reuse	5	5
I2 New opportunities	2 qualified opportunities	8	8
I3 Future risk	1 expected serious incident avoided	7	7
I4 Brand and reputation	20 qualified inquiries	5	5
I5 Talent attraction	10 qualified candidates	5	5
I6 Job satisfaction and culture	100 person-score points	5	5

The weights sum to 100, but 100 is not a cap. Putting 1,200 hours to use yields 30 VP in D1. The model assumes each additional improvement has the same marginal point value within the relevant range. If that assumption does not hold, management must define a shared value function and recalculate all comparisons. A project manager must not unilaterally impose a cap or choose a more favorable curve.

Keep both the weight and the reference fixed. Halving the reference doubles the value of the same raw benefit, changing preferences even if the weight stays the same. The table is a local starting profile, not a claim that 30 customers objectively equal 3,000 hours of continuity. If customer types or incidents are not comparable, segment them and agree on a common calibration before use.

5. Direct value: questions, indicators, and figures

D1. Capacity put to use

Ask: Which tasks take less time? How much active working time is needed before and after? How much review and rework are required? How many relevant tasks use the solution? Which specific work will use the released time, and who allocates it?

Look for: Process logs combined with observations, relevant case volumes, actual usage, and a capacity plan. Shorter elapsed time does not necessarily mean fewer working hours. Compare the same task types and difficulty levels. Include time spent handling errors and exceptions.

Calculate: Hours = relevant volume × usage share × net minutes saved / 60 × realizable share. Training example: 12,000 cases × 80% × 3 minutes / 60 × 62.5% = 300 hours. Net minutes already include review time. The remaining 37.5% of technically saved time cannot be allocated to useful capacity in this project. D1 = 300 / 600 × 15 = **7.5 VP**.

Document and interpret: Save the time sample, selection method, variation, usage measurement, and the manager's plan for the 300 hours. If employees merely “feel faster,” the figure is an assumption. Do not claim cash savings without a separate budget effect. When overtime is actually reduced, assign that specific benefit to D2 and do not count the same hours again here.

D2. Existing expense or investment avoided

Ask: Which existing invoice, capacity purchase, or approved expense will actually disappear? When can the contract be terminated? Would the purchase have been necessary without the project? Who confirms the budget impact?

Look for: Contracts, purchasing plans, termination terms, invoices, and finance approval. An unconfirmed opportunity to buy less is not a certain saving. An employee's unchanged salary, when their time is freed for other work, normally belongs in D1.

Calculate: Three months of avoiding an existing license expense of approximately $1,500 per month gives approximately $4,500 / $22,600 × 15 = **3 VP**. The expense must be avoided within the selected period. Show an avoided one-time investment separately as a one-time effect. Also use a common longer horizon if needed to avoid distorting the comparison.

Document and interpret: Record the gross baseline expense avoided. New project licenses and implementation costs belong in SE and should not also be subtracted in D2. Maintain a separate financial view of actual net savings alongside the points model. Use a consistent price basis across periods; convert D2 spreadsheet inputs to the profile's price year before entering them.

D3. Fewer material errors

Ask: What qualifies as a material error? Who decides? How many relevant cases are there? Is the error rate falling, or is volume simply lower? Have errors shifted to another team?

Look for: An agreed error taxonomy, samples assessed by consistent rules, error rates, and volume. For example, define an invoice error as one requiring correction or causing an incorrect payment. Do not combine minor typos with serious payment errors without a shared assessment rule.

Calculate: 6,000 invoices × (2.0% - 1.6%) = 24 material errors avoided. D3 = 24 / 60 × 10 = **4 VP**. The decrease is 0.4 percentage points, not 0.4 percent. If only some invoices use the solution, limit the volume to that share or calculate each segment separately.

Document and interpret: Save error types, the denominator, assessment instructions, and checks of reviewer agreement. Do not count correction time saved in D1 and again as the full value of D3. There may be two independent effects; if so, explain that D3 represents the remaining harm to the customer or process, with time counted separately. When in doubt, count the benefit once.

D4. Operational continuity

Ask: How many users are actually unable to work during an interruption? For how long? Do they have a viable alternative workflow? Which disruptions does the project affect?

Look for: Incident logs, duration, affected functions, and actual working time lost. Availability percentages are technical indicators; value depends on the business consequences. Do not multiply every employee by the entire outage duration when only a few are blocked.

Calculate: Four fewer interruptions × 1.5 hours × 100 users × 50% actual time lost = 300 user hours protected. D4 = 300 / 3,000 × 10 = **1 VP**.

Document and interpret: Define which incidents the change affects and state the uncertainty around rare events. Remove any hours already counted in D1. Do not count the same avoided incidents again in I3. An absence of outages during a short pilot does not, by itself, demonstrate improved reliability.

D5. Customer acquisition and retention

Ask: Which customers join or stay longer because of the project? What would have happened without it? What pricing, campaign, and market changes are happening at the same time?

Look for: Defined cohorts, CRM status, actual conversion, churn, and comparable periods. Use the same definition of a new customer across projects, such as an active paying customer at the end of the measurement period. Leads are not yet customers.

Calculate: 1,000 comparable customers × (3.0% expected churn without the change - 2.4% with it) = 6 customers retained. D5 = 6 / 30 × 10 = **2 VP**. Acquisition and retention can be added when the groups are distinct and lost customers are also subtracted.

Document and interpret: Save the cohort definition, observation period, and alternative explanations. If the benefit comes from the same customer journey as I4 inquiries, do not reward the entire journey repeatedly. Long contract cycles need later follow-up; an earlier proxy must still be labeled a proxy.

D6. Resource and climate impact

Ask: Which physical resource changes? What system boundary applies? Does consumption increase elsewhere, such as through additional data processing?

Look for: Measured energy, transportation, or material use, plus an identified emissions factor with its year and scope. A green intention or a new sustainability message is not a physical reduction.

Calculate: If the documented net reduction in this example is 2 metric tons CO2e after additional consumption, D6 = 2 / 10 × 5 = **1 VP**. This article does not prescribe a generic emissions factor; the business must document the appropriate factor.

Document and interpret: Retain activity data, the factor, boundaries, and volume adjustments. If another resource, such as water, is strategically important, give it a shared reference in a new profile; do not relabel liters as metric tons CO2e. Brand impact can be counted separately only if it is a distinct documented effect and overlap is addressed.

6. Indirect value: document opportunities without promising the future

I1. Reusable learning

Ask: Which other team can use the learning? What can they do that they could not do before? Has reuse actually been tested? Who takes ownership of the material?

Look for: A specific method or component, documentation, a named recipient, and a test showing that another person or team can use it. A presentation or “we learned a lot” is not validated reuse.

Calculate: Two independent, accepted instances of reuse yield 2 / 4 × 5 = **2.5 VP**. Before implementation, state an expected count with low and high scenarios. Afterward, count only validated reuse.

Document and interpret: Record the learning ID, recipient, application, and acceptance evidence. Repeating the same course for the same purpose is not automatically another instance of reuse. This is an indicator of future value, not hours saved in all future projects. Do not also count the full future benefit from the same learning.

I2. New opportunities

Ask: What previously unavailable outcome does the deliverable make possible? Who needs it? Which barrier has been removed? What does a feasibility test show? What is the next decision?

Look for: An opportunity statement with an owner, a documented need, a successful relevant test, and an agreed next review date. Ten brainstormed ideas are not ten qualified opportunities.

Calculate: One qualified opportunity yields 1 / 2 × 8 = **4 VP**. The count is a rough indicator. When opportunities differ greatly in potential benefit or feasibility, simple counting cannot carry the prioritization decision on its own. Instead, assess the specific platform and its application as a combined investment package.

Document and interpret: Give each opportunity an ID and a link to a future project. Once the application is funded, the portfolio must not count its full future benefit and an unchanged option value as independent benefits. Retain original scores in the history, but address overlap in the current decision overview.

I3. Reduced future risk

Ask: Which future event becomes less likely? What comparable level of severity does it have? What exposure and time horizon apply? What evidence supports the risk reduction?

Look for: Incident history, control tests, exposure, expert assessment, and documented causality. A risk rating moving from “red to yellow” is not directly a numeric probability.

Calculate: For one comparable incident scenario over 90 days: probability without the change 30%, with the change 10%, a difference of 0.20 expected incidents. I3 = 0.20 / 1 × 7 = **1.4 VP**. Model multiple exposures explicitly; do not repeat an aggregate probability for every employee.

Document and interpret: Define incident severity, the assessment basis, sensitivity, and control results. Zero incidents in one period does not prove zero probability. Exclude incidents already covered by D3 or D4. Mandatory safety requirements remain requirements regardless of the score.

I4. Brand and reputation

Ask: Which observable change in reputation is expected to generate relevant interest? How will inquiries caused by that change be distinguished from routine marketing?

Look for: Qualified inquiries with a defined target audience, specific need, contact details, and recorded source. Views, likes, and general awareness are explanatory indicators, not the units selected for VP.

Calculate: Eight additional qualified inquiries, after allowing for the normal trend and overlap with D5 customers, yield 8 / 20 × 5 = **2 VP**.

Document and interpret: Save the qualification rule, channel, baseline, and overlap check. A new sustainability message can influence reputation, but it does not substantiate physical climate benefits or future revenue. If attribution is weak, show a range and a measurement task instead of a precise expectation.

I5. Talent attraction

Ask: Which skills does the business lack? Why would this initiative encourage more relevant candidates to apply? Are additional applicants qualified and above the normal level?

Look for: A consistent qualification rubric, comparable roles, campaign conditions, and recruiting data. Application numbers alone may rise simply because more positions are open.

Calculate: Four additional qualified candidates for comparable hiring needs yield 4 / 10 × 5 = **2 VP**.

Document and interpret: Save the mix of roles, definition of qualified, assessor, and reference period. If recruiting time or agency expenses are benefits under other criteria, define the separate future talent value; otherwise, count the benefit once. Candidates are not the same as hires or demonstrated retention.

I6. Job satisfaction and culture

Ask: Which specific part of daily work improves? Which consistent question measures it? Can we distinguish the change from workload, a new manager, or other simultaneous factors?

Look for: The same measurement instrument before and after, response rates, group composition, and preferably a comparable group. Do not interpret silent employees as satisfied. An anonymous survey can be combined with interviews without implying false precision.

Calculate: The average score of 40 comparable respondents rises by 0.5 points on the agreed 0-10 scale: 40 × 0.5 = 20 person-score points. I6 = 20 / 100 × 5 = **1 VP**. Use either a clearly defined respondent population for both measurements or a justified representative extrapolation. Do not automatically scale the result to all employees.

Document and interpret: For practical purposes, the scale is treated as approximately interval-scaled. The distance between responses is not guaranteed to be objectively equal. This particular weakness must be visible in prioritization. Higher satisfaction also does not prove lower absenteeism or higher productivity. Show measured well-being separately; this profile uses it as an indicator of future cultural value.

7. From scorecards to the total calculation

A training example includes these independent contributions: D1, 300 hours = 7.5 VP; D3, 24 errors = 4 VP; I1, two instances of reuse = 2.5 VP; I2, one opportunity = 4 VP. D4 contributes 450 user hours protected / 3,000 × 10 = 1.5 VP. The direct subtotal is 13 VP and the indirect subtotal is 6.5 VP: **19.5 VP in total**. Protected user hours are a separate operational benefit and are not included in the 300 capacity hours.

Use this calculation to teach transparency: participants must explain all five raw inputs. This is a constructed training example, not a retrospective reconstruction of every criterion in the existing P1 case. When using the spreadsheet, the project's own 12 documented inputs determine the authoritative breakdown.

A portfolio should show central VP estimates, ranges, direct and indirect contributions, evidence limitations, and the largest drivers. VP is a preference-based index with zero defined as no improvement. “20 VP is twice as valuable as 10 VP” holds only within the model's assumptions; it is not a universal property of the business. Proxies and nonlinear effects particularly require judgment.

Calibration as a management exercise

First ask participants to compare specific improvements from zero to each reference point. Which improvement matters most, and why? Then allocate weights and document disagreements. Distinguish strategic preference from evidence: an uncertain benefit needs uncertain inputs, not a hidden reduction in its weight.

Test the profile on at least three different project types, including IT and business changes. Check whether one criterion dominates without a deliberate decision. Test reasonably higher and lower weights while keeping the total at 100. Document which choices change the ranking. Save the approved profile with an owner, date, price year, and scope. This is an EnableAI workflow, not an MCDA literature requirement for a particular number of test projects.

8. Uncertainty and measurement without false confidence

Low, central, and high are plausible scenarios with explanations. They are not automatically the 10th, 50th, and 90th percentiles. Show, for example, how adoption, net time, and actual use of released capacity affect D1. Avoid combining every optimistic assumption in a central estimate. If a probability is already included in an expected improvement, do not apply another general deduction for the same risk.

A sound measurement design compares similar tasks with and without the change. Randomization may be appropriate when practical and responsible. Otherwise, comparable groups, phased rollout, or time series may be better than a single before-and-after measurement. Describe the limitations. The Magenta Book provides a methodological basis for evaluation and attribution. S3: HM Treasury, Magenta Book.

Choose the sample based on variability, desired precision, and the consequences of the decision. “30 observations” is not a universal quality threshold. Report count, selection, missing data, and variation. Rare serious errors need separate controls even when average completion time looks good. For agentic solutions, measure the correct final state and any unauthorized actions, not merely whether the agent says “done.”

9. What can publicly documented cases teach us?

OpenAI's account of Morgan Stanley highlights expert review and ongoing evaluation of an internal knowledge assistant. The training question is: which benefit measures do we still need to calculate our D1? High usage alone does not substantiate 300 released hours. This is a vendor-published case. S4: Morgan Stanley.

Klarna's February 2024 announcement describes both faster case resolution and fewer repeat inquiries. Use it to discuss speed alongside quality. Do not transfer its first-month results to a different organization or treat them as current expectations. This is company reporting, not a controlled EnableAI trial. S5: Klarna.

METR's 2025 study of experienced open-source developers shows why expected and actual productivity should be measured separately. Its result applies to the studied group and setup; it is not a general verdict on AI. Ask participants to explain why satisfaction cannot replace time measurement. S6: METR.

10. How to teach VP to others

Start with the claim “the assistant saves us lots of time.” Have participants formulate questions before seeing a formula. Then provide a dataset covering volume, time, review, adoption, and use of released capacity. Ask them to calculate the raw benefit before VP. Show the weights only at the end. This exposes disagreements about results before they get mixed up with disagreements about priority.

Exercise A, 15 minutes. There are 8,000 relevant tasks, 75% adoption, four minutes of faster processing, one minute of additional review, and 60% usable capacity. What is D1? Answer: 8,000 × 0.75 × 3 / 60 × 0.60 = 180 hours; 180 / 600 × 15 = **4.5 VP**. An answer of 7.5 VP overlooks the realizable share; an answer of 6 VP overlooks review time.

Exercise B, 10 minutes. The team proposes adding “180 × $90 in savings” to D2, even though all employees remain employed and use the time for customer work. Answer: reject the claim of cash savings and the double counting; retain the applied capacity in D1. The SE reference rate does not establish a real D2 saving.

Exercise C, 15 minutes. There are five ideas, two owners, one documented needs test, and no feasibility tests. How many qualified I2 opportunities have been realized? Answer: **0 under the selected definition**. Future potential can be recorded as a justified forecast, but it has not yet met the qualification requirements.

Exercise D, 15 minutes. The target is three net minutes saved per case. The actual result is 2.4, the tolerance is ±25%, and the minimum for continuing is 2.5. Answer: -20%, within the estimation tolerance but below the minimum. Adjust, investigate, or stop; do not automatically scale.

Finish with a short teaching assessment: a participant explains a scorecard of their choice in five minutes, a colleague tries to reproduce the calculation, and the group identifies a possible source of error. Approve the work only when someone else can reproduce the figure and explain its limitations. This is EnableAI's skills assessment, not an external certification.

Sources, further reading, and limitations

S1-S6 are linked beside the relevant arguments above. Further reading: Multi-criteria analysis: a manual, UK Government, 2009. These sources provide methodological foundations and examples. None validates EnableAI's starting weights, 90-day window, or VP/SE as a universal prioritization rule.

Use the question and documentation templates, the SE article, shared follow-up and the five project examples together. The calculation model remains version 3.0. Editorial revision 3.1 expands interpretation and teaching without changing weights or formulas.

Currency: rounded USD equivalents at the ECB reference rates for October 6, 2026. Calculations use unrounded values; displayed amounts may not reproduce exact totals. The original model calibration is retained. Reference rates are resource weights, not US market rates or billing prices.