



Ministry
of Defence

Evaluation Strategy to 2030

A strategy for stronger evidence, better decisions, and better value for money across Defence

Analysis and Performance Directorate

July 2026

1. Introduction

1.1. This strategy updates rather than replaces the Ministry of Defence's evaluation model. It builds on the public strategy published on 26 July 2023, which set out clear expectations that evaluation should focus on the most important MOD interventions, meet defined quality standards, and inform major decisions¹. Those principles remain unchanged. What has evolved is the MOD's level of implementation maturity, alongside a higher benchmark now established through updated cross-government guidance and sustained external scrutiny of evaluation practice.

1.2. MOD has made significant progress in strengthening its evaluation capability and governance over recent years. A dedicated central Evaluation Team is now in place, complemented by specialist evaluation teams embedded within major policy areas including the Office for Veterans' Affairs, People Policy, and Accommodation Policy. Across the investment portfolio, at least 125 Project Evaluation Plans have been endorsed, engagement has taken place with over 250 projects, more than 20 evaluation reports have been delivered, and over 100 analysts have been trained through the Evaluation Academy. Together, these developments represent a strong and credible foundation for evaluation across Defence.

1.3. This strategy therefore marks a deliberate transition to a clearer next phase of implementation. It retains the core governance and principles of the existing approach, while strengthening areas that recent cross-government guidance has brought into sharper focus. It introduces a more explicit theory of change, a clearer model for prioritisation, a defined operating rhythm, stronger expectations around data quality and publication, a more explicit approach to value-for-money judgement, and a set of implementation metrics to track progress over time.

1.4. The strategy also establishes a clear planning horizon, taking MOD towards 2030. This period is intended to be sufficiently long to embed the operating model, apply it consistently to major decisions, and assess whether there has been a material improvement in evaluation coverage, quality, use, and organisational learning; particularly in relation to impact and value for money.

¹ [MOD Evaluation Strategy](#) (Ministry of Defence 2023)

2. Case for Change

2.1. The MOD manages some of government's largest and most complex interventions. These include major equipment programmes, digital and workforce transformation, estate change, policy reform, and cross-cutting capability initiatives. They are high value, politically and operationally sensitive, and often delivered over long timescales. In that context, weak evaluation does not only create an analytical gap. It creates a decision-making gap.

2.2. That challenge is especially acute for Defence major projects, where hard-to-measure outcomes and long delivery horizons make it harder to set baselines early, sustain data collection, and generate timely evidence on impact and value for money.

2.3. External scrutiny has repeatedly made the same underlying point. Evaluation is a value-for-money issue, not an optional add-on². Parliamentary scrutiny has linked poor use of evaluation and modelling to wider capability and accountability concerns across government³. In defence specifically, parliamentary and committee reporting has pointed to system-level weaknesses in procurement, delivery, and learning rather than isolated project failure⁴.

2.4. The 2023 MOD strategy addressed this by setting a clear governance direction: increase evaluation coverage, improve quality and impact, and ensure evaluation informs major decisions⁵. That was and remains the right foundation. The Department's own progress since then suggests the model is working. Evaluation is more visible in business case development, more firmly embedded in the central scrutiny applied at approval and review points and better supported through capability-building and theory of change development by the central evaluation team, wider Analysis Directorate, and evaluation professionals embedded in policy areas.

2.5. This update to the strategy therefore keeps the governance core of the existing strategy while setting out more clearly what MOD is trying to achieve, how the system should operate, and how progress should be judged over time.

² [Evaluating Government Spending](#) (National Audit Office, 2021)

³ [Use of Evaluation and Modelling in Government](#) (Committee of Public Accounts, 2022)

⁴ [It's Broke – and it's Time to Fix It: The UK's Defence Procurement System](#) (Defence Committee, 2023)

⁵ [MOD Evaluation Strategy](#) (Ministry of Defence, 2023)

3. Vision, Outcomes and Strategic Objectives

3.1. Vision

MOD will plan, deliver, and use evaluation as a routine part of how Defence manages risk, improves delivery, and demonstrates value for money. Evaluation will begin early, be proportionate to the scale and uncertainty of the intervention and be used systematically to inform Defence decisions.

3.2. Intended Outcomes

By the end of the strategy period, MOD should have:

1. Stronger evaluation coverage of the most important projects, programmes, and policies.
2. More interventions with evaluation considered early in design rather than late in delivery.
3. Higher-quality evaluation design, including stronger theories of change, clearer questions, and better data planning.
4. More proportionate evaluation plans across centrally approved projects.
5. Stronger evaluation coverage of the highest-priority centrally approved projects, in particular the Government Major Projects Portfolio and the Department Major Projects Portfolio, and of other priority projects drawn from the Defence Plan.
6. More consistent use of evaluation evidence in business cases, approvals, portfolio reviews, and strategic planning.
7. More credible and transparent value-for-money assessment, including wider use of rubric-based judgement (where performance is assessed against clear criteria and standards).
8. Stronger evaluation capability among analysts, senior responsible owners, programme teams, and commissioners.
9. Internal learning enabled through an evaluation evidence database, including external evaluations, informing Defence decisions.
10. More publication and synthesis where security and sensitivity rules allow.
11. Better evidenced Defence decisions, earlier delivery adaptation, and stronger accountability for results.

3.3. Strategic Objectives

The Department will pursue five strategic evaluation objectives:

1. **Embed evaluation in the Defence project lifecycle.** Clarify, update, and mandate the Defence evaluation operating model so that clear, proportionate evaluation expectations are defined and applied at each stage of the Defence project lifecycle, and evaluation is embedded within routine management, assurance, and approval processes. These processes should be clearly aligned to Government Major Projects Portfolio and Treasury Approval Processes on Evaluation.
2. **Enhance the quality and relevance of Defence evaluations.** Strengthen the rigour of Defence evaluations by building on strong process evaluation foundations and applying fit-for-purpose designs. Ensure evaluation methods are proportionate, contextually appropriate, and explicitly aligned to decisions,

so evaluations can generate credible evidence on impact and value for money.

3. **Strengthen the use of evidence in Defence decision-making.** Ensure evaluation evidence routinely informs business cases, approvals, portfolio reviews, delivery adaptation, and future strategy, supported by an accessible and searchable evidence base of relevant internal and external evaluations.
4. **Build internal capability and supporting infrastructure.** Equip analysts, senior responsible owners, delivery teams, commissioners, data owners, and governance bodies with the skills, tools, and support needed to commission, deliver, interpret, and use evaluation effectively, including theory-based, quasi-experimental, and experimental approaches where appropriate and proportionate.
5. **Review and improve the evaluation system over time.** Monitor whether the strategy is improving evaluation coverage, quality, use, transparency, and learning across Defence, and use this evidence to refresh the operating model, guidance, and capability offer over time (performance measures in Section 11).

4. Theory of Change

4.1. The core problems are that evaluation in Defence has historically been rare, has not consistently followed appropriate guidance and standards, and has not been well connected to decision-making. When interventions do not start with clear outcome logic, data plans, and evaluation questions, MOD loses the chance to learn credibly from major investments and reforms. The strategy controls some of this directly, such as governance, guidance, review, and capability-building. It influences other outcomes indirectly, including whether senior leaders use evidence well and whether programmes maintain good data over time. The strategy cannot remove all external constraints, but it can materially improve whether MOD enters major decisions with better evidence than it does today.

4.2. This top-level theory of change sets out how Defence's evaluation system functions as a reinforcing mechanism, with each component linked through a causal chain. The process moves from inputs to activities, outputs, outcomes, and ultimately to the long-term impact.

Table 1: Theory of Change

► Inputs	► Activities	► Outputs	► Outcomes
- Senior leaders and strategy owners - Central Evaluation Team and specialist support - Governance forums, approval gateways, and escalation routes - Data, evidence infrastructure, and curation capacity - Analyst teams, training materials, and business-area teams - Methods guidance, Evaluation Hub, and standards	- Running the operating model - Applying quality and VfM standards - Embedding evidence in decision processes - Delivering capability, support, and infrastructure - Reviewing performance and adapting the system	- Lifecycle rules, tiering decisions, and plan requirements - Scoping packs, endorsed plans, theories of change, designs, and VfM frameworks - Decision-use products - Training delivered, teach-ins, labs, specialist support, and Hub tools - Portfolio reporting, review insights, refreshed evidence - Completed evaluation reports	- Earlier and more consistent evaluation - More credible and useful evaluations and reports - Evidence easier to find and used more consistently - Resources and scrutiny focused on highest-value gaps

4.1. Inputs

Six critical inputs underpin the system. Senior leaders and strategy owners establish expectations, demand quality, and create the environment for evaluation to shape decision-making. The Central Evaluation Team and specialist support offer design support, review, capability-building, and act as stewards of the system. Governance forums, approval gateways, and escalation routes give the operating model its authority, ensuring evaluation requirements are enforced. Data, evidence infrastructure, and curation capacity are essential for monitoring, synthesis,

Registry/Hub curation, and maintaining databases. Analyst teams, training materials, and business-area teams are responsible for delivering evaluation work and embedding the model across Defence. Finally, methods guidance, the Evaluation Hub, and standards ensure consistency and raise the quality of all evaluation activities. Together, these elements form the resource base that drives the system's delivery.

4.2. Activities

These inputs enable five key strategic activities. Running the operating model involves integrating evaluation across the project lifecycle, applying tiering, and undertaking early scoping. Applying quality and value-for-money standards ensures that designs, baselines, theories of change, criteria, and frameworks meet minimum credible standards. Embedding evidence in decision processes means that business cases, reviews, approvals, and portfolio decisions are informed by evaluation findings. Delivering capability, support, and infrastructure is achieved through training, specialist support, updates to the Evaluation Hub, and database curation. Reviewing performance and adapting the system involves using portfolio data, quarterly reviews, and annual syntheses to update priorities and continually improve the model. Collectively, these activities drive the generation and use of high-quality evaluation evidence within Defence.

4.3. Outputs

The activities described produce six main outputs. Lifecycle rules, tiering decisions, and plan requirements are generated through the operating model. Scoping packs, endorsed plans, theories of change, endorsed designs, and frameworks result from the application of quality standards. Decision-use products, such as briefings, syntheses, and access routes, emerge from embedding evidence in decision-making. Training sessions, teach-ins, labs, specialist support, and Hub tools are delivered through capability and infrastructure work. Portfolio reporting, review insights, and refreshed evidence arise from system review and adaptation. Finally, completed evaluation reports, which are registered and published where possible, are produced through a combination of quality assurance and evidence-use processes. The system is designed not only to set expectations and build infrastructure but also to generate completed, usable evaluation evidence.

4.4. Outcomes

Collectively, these outputs strengthen system behaviour in four important ways. First, they enable earlier and more consistent evaluation, supported by clear lifecycle rules and ongoing training and support. Second, they lead to more credible and useful evaluations and reports, driven by robust designs, strong capability support, and especially the production of completed evaluation reports. Third, they make evidence easier to find and use more consistently, with decision-use products and the publication or registration of reports playing a key role. Finally, they help shift resources and scrutiny towards the highest-value gaps, informed by portfolio reporting and outputs from system reviews. These outcomes demonstrate how the operating model improves the timing, quality, accessibility, and strategic targeting of evaluation across Defence.

4.5. Impact

Ultimately, all outcomes contribute to the overarching impact: better evidenced decisions, value for money, learning, and accountability. This is achieved through

earlier evaluations, more credible reports, more consistent evidence use, and strengthened prioritisation. The impact reflects a mature system where evaluation is routinely embedded, credible, and targeted where it is most needed.

4.6. Risks and Assumptions

4.6.1. The following assumptions sit within the theory of change and describe the causal conditions required for delivery activity to translate into improved use of evaluation in decision-making. Each assumption is therefore also a risk driver. The strategy does not treat these as static background conditions; instead, they will be actively monitored and managed⁶ as part of delivery.

4.6.2. The key assumptions are that:

1. There are sufficient resources (funding, time, and analytical capacity) to deliver this strategy.
2. There is sufficient evaluation capability across Defence to deliver this strategy.
3. Senior leaders continue to demand and use evidence in major decisions.
4. Programmes cooperate with evaluation requirements and maintain engagement over time.
5. Data can be accessed securely and in forms usable for monitoring and evaluation.
6. Operational tempo does not crowd evaluation out during delivery.
7. Rubric criteria and standards are accepted as credible by decision-makers and delivery teams.
8. Relevant external evaluations can be identified, curated, and used appropriately in Defence contexts.
9. Methods remain proportionate and feasible within security and practical constraints.

4.6.3. The strategy's performance framework will follow the same logic chain and will be used to test whether these assumptions continue to hold. This requires a clear distinction between:

- **enabling conditions** (the assumptions above),
- **delivery actions** (what the strategy implements),
- **immediate products** (such as plans, guidance, and completed evaluations),
- **behavioural change** (how evidence is demanded and used), and
- **risks to the causal pathway** (where assumptions begin to weaken).

4.6.4. By tracking both delivery progress and the health of these assumptions, the central evaluation team within MOD will be able to intervene early, for example through targeted engagement, changes to guidance, or senior escalation, where risks emerge. Without this approach, there is a material risk that the Department treats implementation activity as evidence of impact, rather than actively managing whether the strategy is genuinely improving decision making and value for money.

⁶ Active management means that the department will: (1) make these assumptions and associated risks explicit; (2) monitor whether they are holding in practice; (3) take corrective action where they weaken; and; (4) escalate material risks where they threaten the strategy's causal pathway.

5. Scope and Prioritisation

5.1. Scope

5.1.1. This strategy applies to interventions where evaluation can most improve Defence decision-making and learning. In line with government requirements, all programmes and projects that are Government Major Projects or meet the Transforming Approvals Process (TAP) criteria are required to have robust evaluations⁷. Other programmes and projects should also be evaluated in a robust and proportionate manner, consistent with the expectations set out in the Magenta Book. Accordingly, this strategy covers:

1. Centrally approved projects and programmes, with particular emphasis on those in the Government Major Projects Portfolio, the Department Major Projects Portfolio, or those meeting TAP criteria.
2. Critical policies linked to major departmental delivery or spending decisions.
3. Major cross-cutting reforms, including workforce, digital, commercial, estate, and operating-model change.
4. Novel pilots or demonstrators where uncertainty is high and learning value is likely to extend beyond a single team.

5.1.2. The strategy does not require every piece of routine business-as-usual activity to have a full evaluation. Proportionality remains essential. It also does not replace audit, assurance, test and evaluation, requirement management, nor performance reporting. Instead, it should align with these processes, use their evidence where appropriate, and avoid duplication.

5.2. Tiered Evaluation Approach

5.2.1. Consistent with the HM Treasury *Magenta Book*, the Government seeks to undertake evaluations that are proportionate, fit for purpose, and genuinely useful to decision makers. Adopting a tiered evaluation approach mitigates two common risks: firstly, the risk of deploying highly resource-intensive evaluation methods where the value for decision making is low; and secondly, the risk of relying on minimal evidence arrangements in cases where levels of spend, risk, or uncertainty warrant more robust evaluation.

5.2.2. Proportionality means tailoring evaluation effort to decision risk and evidence gaps, not layering additional assurance where strong governance, performance management, or technical assurance already exist. Evaluation should draw on existing evidence wherever possible and should not duplicate established delivery or assurance mechanisms unless there is a clear decision need.

5.2.3. MOD should therefore assign each intervention to one of three evaluation tiers at the scoping stage, with the assigned tier reviewed if there are material changes in spend, risk, delivery performance, or strategic importance.

⁷ [Treasury Approvals Process for Projects and Programmes](#) (HM Treasury, 2026)

Table 2: Tiered Evaluation Approach

Tier	Typical use case	Typical Evaluation Activities	Minimum expectation for evidence and evaluative thinking
Tier 1: Enhanced evaluation	Major, novel, high-risk, high-cost, or highly scrutinised interventions where decisions have significant fiscal, operational, or strategic consequences	High level of methodological rigour. Evaluation should form part of early policy design, programmes should embed strong governance, more intensive data collection, and the most credible feasible causal and value-for-money methodologies	Senior oversight, early scoping before the main approval route, robust theory of change, defended methodological choice, funded data collection, formal quality assurance, and a clear plan for synthesis and decision use
Tier 2: Standard evaluation	Interventions of moderate spend, strategic importance, uncertainty, or learning value, where MOD needs stronger evidence on implementation and outcomes but a major bespoke study would be disproportionate	Structured process and impact evaluation, clearer baselines, planned data collection, and scheduled reporting into approvals and portfolio review	Endorsed evaluation plan, explicit questions, feasible methods, baseline arrangements, implementation learning, and outcome reporting timed to key decisions
Tier 3: Monitoring and light-touch evaluation	Lower-risk ⁸ or lower-value interventions, repeat activities, or areas where existing evidence is already strong and the main need is implementation visibility	Limited additional evaluation effort beyond good monitoring, process evaluation, and use of existing evidence	A proportionate evaluation plan, clear theory of change, identified data owner, use of relevant prior evidence, and a decision on whether any further evaluation is needed later Tier 3 arrangements are intended to rely primarily on existing monitoring, assurance, and evidence, with evaluation activity focused on sense-checking assumptions and supporting portfolio visibility rather than generating new burdens.

⁸ For the purposes of tiering, “risk” should be assessed using existing Defence approval evidence rather than a separate evaluation-specific judgement. This includes information already generated through the approval process, such as programme and project risk registers, Risk and Complexity Assessments, and the assigned approval route (for example, TAP category, GMPP status, or delegated approval). An intervention may be considered “lower risk” where approval evidence indicates lower levels of uncertainty, delivery complexity, decision consequence, and value-for-money exposure, rather than lower importance. Tiering judgments should be reviewed if approval evidence changes materially over time.

5.2.4. These tiers are about ensuring MOD conducts proportionate, fit-for-purpose evaluations that are genuinely useful to decision makers, not prestige. A Tier 3 intervention is not unimportant; it is an intervention where the most proportionate response is disciplined monitoring, use of existing evidence, and a lighter evaluative burden. Equally, not every Tier 1 intervention will allow for experimental or quasi-experimental methods due to the lack of appropriate counterfactual. The test is whether the level of effort matches the value of reducing uncertainty for Defence decisions.

5.2.5. In practice, centrally approved projects should normally be expected to meet at least Tier 3 requirements, priority projects and policies will often need Tier 2 arrangements, and the most material major projects, major reforms, and novel capability areas should be considered for Tier 1 from the outset. The Central Evaluation Team should maintain the tier assignment in the forward plan so that leadership can see both evaluation coverage and the intensity of support being committed across the portfolio.

6. Operating Model

Evaluation should be built into the normal management cycle of Defence, not treated as a parallel analytical exercise. The operating model has two dimensions: lifecycle integration for individual interventions and portfolio review across the department.

6.1. Lifecycle Expectations

These lifecycle expectations should be read alongside GovS 002⁹, which requires a defined life cycle with approval gates, decision points, business cases, and assurance reviews, and alongside the current NISTA gateway sequence set out in the government project assurance framework as Gate Reviews 0 to 5¹⁰. For major projects, evaluation milestones should be reflected in the integrated assurance and approval plan so that scoping, baseline readiness, and findings are available before the relevant approval and gateway decisions. These expectations will be enforced through existing approval, assurance, and portfolio governance routes, rather than through new standalone processes. Repeated failure to meet proportionate evaluation expectations will be escalated through investment approvals and analysis function governance.

Table 3: Lifecycle Expectations

Stage	Evaluation expectation	GovS 002 / gateway alignment
Strategic Outline Case	Define intended outcomes, measures of success, major assumptions, an initial theory of change, and any likely value-for-money criteria. Identify how much evaluation effort the intervention is likely to need and what existing evaluation evidence is already relevant.	Corresponds to the front end of the defined project life cycle. For programmes and the earliest strategic framing, this aligns most closely to Gate 0: Strategic Assessment .
Outline Business Case	Complete evaluation scoping. Define the main questions, likely methods, baseline needs, data constraints, decision points the evaluation should inform, and the existing evaluation evidence that should be referenced. Decide need for evaluation of assessment phase.	Should support early approval and assurance activity. This aligns most closely to Gate 1: Business Justification and, where delivery options and the delivery model are being settled, Gate 2: Delivery Strategy .
Full Business Case	Secure an endorsed evaluation plan for interventions in scope. Confirm governance, delivery responsibilities, quality assurance, funded data collection arrangements, and the value-for-money approach where required. Report results of assessment phase evaluation as appropriate.	The evaluation plan should be settled before major commitment of funding and contract award. This aligns most closely to Gate 3: Investment Decision .

⁹ [Government Functional Standard GOVS002: Project Delivery](#) (Government Project Delivery and Cabinet Office, 2025)

¹⁰ [Government Project Delivery Assurance Framework](#) (Infrastructure and Project Authority, 2023)

Stage	Evaluation expectation	GovS 002 / gateway alignment
Early delivery and implementation	Run process evaluation and implementation learning so that early delivery problems, context shifts, and data issues can be identified in time to act on them.	Takes place after Gate 3 and before Gate 4 , so mobilisation risks, data quality, and implementation issues can inform readiness decisions.
Readiness for service and initial operating capability	Confirm that monitoring, data flows, and evaluation governance work in live conditions, and use early evidence to test whether intended users, processes, and benefits measures are functioning as expected.	Aligns most closely to Gate 4: Readiness for Service.
Full operating capability and benefits realisation	Deliver the main impact, and value-for-money work that is feasible for the intervention. Feed findings into portfolio review and future investment decisions.	Aligns most closely to Gate 5: Operations Review and Benefits Realisation.
Closure, refresh, or follow-on investment	Produce synthesis and lessons learned. Ensure findings shape successor activity, policy refresh, procurement, or wider departmental practice. Publish final evaluation reports to allow findings to be used for future decision making.	Use the synthesis from Gate 5 and later in-service evidence to inform closure decisions, refreshes, and any successor Gate 0 / Gate 1 work for follow-on investment.

6.2. Portfolio Rhythm

6.2.1. At departmental level, MOD should operate a recurring evaluation cycle.

1. Maintain a forward evaluation plan covering the interventions where evaluation is most needed for decision-making, learning, and value for money.
2. Review coverage, quality, and delivery progress at least quarterly through the relevant portfolio and analytical governance routes.
3. Refresh priorities annually to align with strategic planning, affordability, and delivery risk.
4. Maintain and refresh a searchable database of internal and external evaluations relevant to Defence interventions.
5. Produce an annual synthesis for senior leaders summarising completed evaluations, emerging findings, major gaps, and implications for the portfolio.

6.2.2. This operating model reflects the lessons in recent cross-government guidance: evaluation works best when it is visibly connected to policy development, monitoring, approvals, planning cadence, and portfolio review rather than standing alone.

6.2.3. For Defence portfolios with long delivery horizons, evaluation evidence is not expected to provide definitive answers within the lifecycle of individual programmes. Its primary role is to support portfolio-level strategic judgement over time by testing assumptions, identifying emerging patterns, and informing choices about balance, sequencing, adaptation, and future investment beyond the scope of any single programme.

6.2.4. Portfolio-level synthesis should focus on thematic learning across groups of programmes, drawing together evidence from multiple evaluations to identify recurring delivery patterns, common risk drivers, and transferable lessons. This comparative perspective is intended to inform strategic portfolio choices, rather than to re-litigate individual programme performance. This ensures that evaluation effort generates insight at a scale that is meaningful for portfolio leadership, rather than relying on isolated findings from individual studies.

6.2.5. Taken together, this portfolio rhythm is designed to maximise learning and strategic insight while minimising unnecessary burden on delivery teams, by focusing evaluation effort where it adds the greatest decision value and by synthesising evidence at portfolio level rather than expanding assurance at programme level.

7. Design Standards and Methods

7.1. This strategy does not prescribe a single evaluation method for every intervention. It sets minimum expectations for quality and proportionality in line with the Magenta Book¹¹.

7.2. All in-scope evaluations should:

1. Start from an explicit theory of change.
2. Define a small set of clear and decision-relevant evaluation questions.
3. Use methods that are proportionate to the maturity, scale, complexity, and feasibility of the intervention.
4. Explain how monitoring data, management information, and other operational evidence will support the evaluation.
5. Address value for money where the intervention is material enough for that judgement to matter.
6. Use explicit evaluative criteria and performance standards where an overall value-for-money judgement is required, so that economy, efficiency, effectiveness, equity, and wider public value trade-offs are assessed transparently.
7. Meet appropriate standards on quality assurance, ethics, security, and transparency.

7.3. MOD should expect a mixed-methods approach. Depending on the context, this may include process evaluation, experimental and quasi-experimental impact designs, theory-based approaches (e.g. contribution analysis and realist evaluation), value-for-money assessment, and synthesis across multiple evidence sources.

7.4. For value-for-money work in particular, MOD should use a rubric-based approach where a unit-cost metric would be too narrow. That means defining the criteria for judgement in advance, specifying what stronger and weaker performance look like, and then testing the available evidence against those^{12 13}. This is especially important for Defence interventions that are high-cost, multi-objective, operationally sensitive, or expected to deliver value across effectiveness, learning, resilience, and strategic fit rather than through one measurable outcome alone.

7.5. The practical standard is not methodological maximalism. It is proportionality. Some Defence interventions will never be suited to clean counterfactual designs. That does not remove the need for evaluation. It increases the importance of strong theories of change, clear baselines, transparent assumptions, and disciplined use of multiple evidence sources.

¹¹ [The Magenta Book](#) (HM Treasury and Evaluation Task Force, 2026)

¹² [Assessing Value for Money](#) (King, et al., 2023)

¹³ [A Guide to Evaluation of Value for Money in UK Public Services](#) (King & Hurrell, 2024)

8. Capability, Roles and Resources

8.1. Evaluation capability in MOD must extend beyond specialists alone. A credible system depends on clear responsibilities across central functions, analysts, delivery teams, and senior leaders.

Table 4: Roles and Responsibilities

Role	Core responsibility
Central Evaluation Team	Set standards, advise on prioritisation, review plans, support major evaluations, maintain portfolio visibility, and lead capability-building
Programme and project SROs ¹⁴	Own evaluation scoping and delivery for interventions in scope, ensure proportionate resourcing across the lifecycle, and ensure findings are used and transparently reported.
Analysts and evaluation leads in business areas	Develop theories of change, design proportionate and robust evaluations, manage delivery, assure quality, and interpret findings for decision-makers
Policy, finance, commercial, and data owners	Ensure the intervention remains evaluable through good design, baseline creation, contract provisions, and data access arrangements
MOD Investment Approval Committee	Review coverage, challenge quality, resolve escalation, and ensure evidence is used in major decisions

8.2. Evaluation design and delivery should be undertaken by, or in active consultation with, staff who have appropriate evaluation expertise. This includes engagement with relevant government profession specialists (for example economists, social researchers, operational researchers, statisticians) to ensure that methods are credible, proportionate, and aligned with government evaluation standards.

8.3. MOD should build on the capability investments already made through the Evaluation Academy. The next phase should expand from analyst training alone to a broader operating model that also equips SROs, project delivery professionals, and commissioners to ask for and use good evaluation, including rubric-based value-for-money judgement where appropriate.

8.4. That capability model should include a specialist methods offer. The Evaluation Academy should add structured modules, practical workshops, and design clinics covering experimental, quasi-experimental, and theory-based evaluation design. MOD should also maintain a small cross-Defence network of method leads who can advise on scoping, support peer review of high-stakes designs, and help teams choose feasible approaches early enough for baselines, comparators, and data collection to be put in place.

¹⁴ [Role of senior responsible owner](#) (Government Project Delivery Profession, 2025)

8.5. For the most important evaluations, capability-building should happen through delivery as well as training. Priority projects should pair business-area analysts with central evaluation specialists, and any external methodological support should include explicit knowledge transfer so that design capability is retained inside Defence rather than bought in afresh each time.

8.6. The Department should also maintain a clear resourcing model. Priority interventions should identify evaluation ownership, delivery resource, and data collection requirements at approval stage rather than relying on later ad hoc funding.

9. Governance and Accountability

9.1. Governance must make evaluation both visible and consequential. The aim is not to create a parallel bureaucracy, but to ensure that evaluation requirements are taken seriously at the same points where MOD already makes major decisions.

9.2. The governance model should include:

1. Director of Analysis & Performance, as the owner for the strategy, with the Analysis and Performance Directorate Evaluation Team as the central function responsible for coordinating delivery.
2. Clear approval expectations for priority interventions and other projects or programmes requiring endorsed evaluation plans, with the key mechanisms being Joint Service Publication 507: Investment Appraisal and Evaluation and Joint Service Publication 655: Investment Approvals.
3. Regular reporting on evaluation coverage, quality, slippage, and key findings through the MOD Investment Approval Committee and related Defence governance routes.
4. Escalation where interventions repeatedly fail to produce adequate scoping, data planning, or evaluation delivery.
5. Annual review of the forward evaluation plan by the MOD Investment Approval Committee.

9.3. This continues the strongest feature of the 2023 MOD strategy: evaluation should be visible to senior governance and should shape major portfolio choices. The difference in this strategy is that governance is now tied more clearly to lifecycle expectations, portfolio review, and implementation metrics.

10. Learning, Transparency and Dissemination

10.1. Evaluation should improve Defence performance, not just produce reports. MOD should adopt the following principles:

1. Register planned and completed evaluations on the Evaluation Registry where cross-government rules apply, with publication by default where classification and sensitivity rules permit¹⁵.
2. Follow the Government Social Research publication protocol for planning, release, and communication of evaluation outputs¹⁶.
3. Maintain a searchable internal database of evaluations and synthesis that covers MOD work and external evaluations relevant to Defence decisions, with clear metadata and proportionate quality signals.
4. Use that database to improve on a narrow commitment to cite evidence in business cases by expecting new business cases to reference the most relevant evaluation evidence available.
5. Require management response or follow-up action for material evaluations so that findings are acted on rather than simply noted.
6. Produce periodic cross-portfolio syntheses to identify recurring delivery patterns, evidence gaps, and reusable lessons.

10.2. Publication in a defence context will never be unrestricted. But the default should still be to publish where feasible and to circulate findings internally where external release is not possible.

10.3. As of 18 May 2026, the Evaluation Registry search filtered to Ministry of Defence returns 33 entries: 26 planned, 4 ongoing, and 3 complete. That provides a practical baseline for tracking whether MOD is improving registration discipline, moving evaluations through delivery, and publishing completed work where security and sensitivity rules permit.

¹⁵ [Guidance on using the Evaluation Registry](#) (Evaluation Task Force, Cabinet Office, and HM Treasury, 2024)

¹⁶ [Government Social Research: Publication Protocol](#) (Civil Service and Government Social Research Profession, 2022)

11. Implementation, Metrics and Review

This strategy sets a planning horizon towards 2030 and should be implemented in phases.

11.1. Phase 1: Reset And Baseline

During the first year, MOD should confirm delivery arrangements under the strategy owner, Director of Analysis & Performance, refresh the forward evaluation plan, confirm prioritisation criteria and decision rules, and set baselines for the performance measures below. This phase should also close the most obvious gaps in scoping, theory of change quality, data planning, publication processes, value-for-money criteria development, and evaluation database design.

11.2. Phase 2: Embed in Routine Business

During the second year, evaluation expectations should be fully aligned with approval points, quarterly portfolio reporting, and annual strategic review. Demonstrator evaluations should be used to test the model in high-priority areas, including rubric-based value-for-money assessment. This phase should also establish the specialist methods offer in experimental, quasi-experimental, and theory-based design, using demonstrator projects, design peer review, and targeted support to build internal capability through live delivery. The evaluation database should be populated and used to support business case development and portfolio review.

11.3. Phase 3: Improve and Refresh

During the third year, MOD should review whether the strategy has materially improved coverage, quality, use, and learning. That review should inform the next strategy refresh and any decision to split the model more formally into a public strategy, an operating annex, and a rolling portfolio plan.

11.4. Performance Measures

11.4.1. This table provides an overview of the key performance measures linked to the MOD's evaluation strategy. It details the ambitions for each metric, including coverage of centrally approved interventions and priority major projects, as well as plans for enhancing capability and developing evidence-based tools. The focus is on future improvements, emphasising ongoing progress and the targets set for development. The table provides a clear view of progress to date and the targets established to enhance evaluation quality, reach, and capability building through training and specialist support.

Table 5: Performance Measures

Measure	Ambition
Coverage of centrally approved interventions with endorsed evaluation plans at the main approval stage	Increase year on year until proportionate evaluation plans are in place for all centrally approved projects
Coverage of priority major projects with active plans	Full coverage (100%) of priority major projects with active plans and regular reporting on delivery
Capability built through training and specialist support	Maintain analyst pipeline, extend capability expectations to SROs, delivery teams, and commissioners, and establish a specialist methods offer covering experimental, quasi-experimental, and theory-based design
Evidence base and value-for-money tools produced	Searchable evidence base maintained, and rubric-based value-for-money frameworks in place for material interventions
Transparency and dissemination products	Registry compliance and publication by default for releasable evaluation outputs, with growth in completed and published evaluations over time
Completed evaluations	Sustained pipeline of completed evaluations, with portfolio synthesis used to interpret findings and gaps
Use in decision-making	Evidence from completed or existing evaluations routinely cited in business cases, portfolio reviews, delivery decisions, and strategy refreshes

11.4.2. The strategy should be formally reviewed in 2030, with annual progress reporting in the meantime. The review should test not only whether inputs and activities were delivered, but whether the strategy improved output quality, decision-use, portfolio visibility, risk management, and the quality of evaluation practice across Defence over the review period.

12. Conclusion

12.1. MOD already has the beginnings of a credible departmental evaluation system. The next step is not to abandon the current model, but to make it more explicit, more operable, and more measurable. This draft does that by keeping the core strengths of the July 2023 strategy while adding the theory of change, operating model, learning arrangements, rubric-based value-for-money practice, and implementation discipline expected of a stronger departmental strategy today.

12.2. If implemented as intended, the strategy will give MOD a clearer basis for deciding what to evaluate, when to evaluate it, how to govern evaluation, and how to use findings to improve Defence outcomes and value for money.

13. References

- Civil Service and Government Social Research Profession. (2022). *Government Social Research: Publication Protocol*. Retrieved from <https://www.gov.uk/government/publications/government-social-research-publication-protocols>
- Committee of Public Accounts. (2022). *Use of Evaluation and Modelling in Government*. Retrieved from <https://publications.parliament.uk/pa/cm5803/cmselect/cmpubacc/254/report.html>
- Defence Committee. (2023). *It Is Broke — and It's Time to Fix It: The UK's Defence Procurement System*. Retrieved from <https://publications.parliament.uk/pa/cm5803/cmselect/cmdfence/1099/report.html>
- Evaluation Task Force, Cabinet Office, and HM Treasury. (2024). *Guidance on using the Evaluation Registry*. Retrieved from <https://www.gov.uk/guidance/guidance-on-using-the-evaluation-registry>
- Government Project Delivery and Cabinet Office. (2025). *Government Functional Standard GovS 002: Project Delivery*. Retrieved from <https://www.gov.uk/government/publications/project-delivery-functional-standard>
- Government Project Delivery Profession. (2025). *Role of the senior responsible owner*. Retrieved from <https://projectdelivery.gov.uk/library-product/the-role-of-the-senior-responsible-owner/>
- HM Treasury. (2026). *Treasury Approvals Process for Projects and Programmes*. Retrieved from <https://www.gov.uk/government/publications/treasury-approvals-process-for-programmes-and-projects>
- HM Treasury and Evaluation Task Force. (2026). *Magenta Book*. Retrieved from <https://www.gov.uk/government/publications/the-magenta-book>
- Infrastructure and Project Authority. (2023). *Government Project Delivery Assurance Framework*. Retrieved from <https://projectdelivery.gov.uk/library-product/government-project-assurance-framework/>
- King, J., Wate, D., Namukasa, E., Hurrell, A., Hansford, F., Ward, P., & Faramarzifar, S. (2023). *Assessing Value for Money: the Oxford Policy Management Approach Second Edition*. Retrieved from <https://www.opml.co.uk/sites/default/files/2024-06/opm-value-money-vfm-approach-v2-1.pdf>
- Ministry of Defence. (2023). *Evaluation Strategy*. Retrieved from <https://www.gov.uk/government/publications/ministry-of-defence-evaluation-strategy>
- National Audit Office. (2021). *Evaluating Government Spending*. Retrieved from <https://www.nao.org.uk/reports/evaluating-government-spending/>
- National Audit Office. (2022). *Evaluating Government Spending: An Audit Framework*. Retrieved from <https://www.nao.org.uk/insights/evaluating-government-spending-an-audit-framework/>