

Evidence: Delivering value from government investment in major programmes

March 2024

This is the [Centre for Public Data](#)'s evidence to the Public Accounts Committee's [inquiry](#) on delivering value from government investment in major programmes.

Written evidence

The problem: we lack the data to evaluate value for money

1. Delivering value for money depends on metrics. Metrics tell us if the intended outcomes are being delivered and if things are going wrong. They also help us learn from what has gone right and wrong in the past.
2. The Government can only evaluate the value for money of major programmes if it has the required metrics. In theory, HM Treasury guidance is that major government programmes and projects should define and monitor success metrics.
3. But this is often not the case. In 2019, the Committee [found that](#) only 8% of £432 billion of major project spending had robust impact evaluation plans in place, and 64% of spend had no impact evaluation arrangements.

Missing metrics in major programmes: an example

4. An example of missing metrics making it hard to assess the value of a major programme, and potentially leading to wasted spending, is the National Tutoring Programme (NTP). The NTP was launched by the Department for Education (DfE) in 2020 and is ongoing.
5. The DfE Accounting Officer's [assessment](#) of the NTP states that its key objective is to "increase the *number of disadvantaged children receiving high quality tutoring* in order to address the impact of Covid-19 on the *educational outcomes of disadvantaged pupils*" (emphases ours).
6. To monitor if NTP delivered value for money by meeting its objective, DfE would therefore to know:
 - a. the number of disadvantaged pupils receiving tutoring
 - b. the quality of the tutoring they received
 - c. the impact on their educational outcomes.

7. However, DfE did not apparently monitor any of these metrics initially:
 - a. **Disadvantaged pupils being tutored:** In early 2022, the DfE [told Parliament](#) it could not provide breakdowns of pupil numbers because it “does not collect data on pupils” (only the count of pupils per school, via the school census). The first count of disadvantaged pupils receiving tutoring was not published till [autumn 2022](#).
 - b. **Quality of tutoring:** In early 2022, DfE also [told Parliament](#) it did not record how many NTP tutors were qualified teachers, though concerns were raised about tutoring quality soon after the NTP programme began.
 - c. **Outcomes:** While outcomes naturally take longer to measure, again the first outcomes were not published until autumn 2022. These found little evidence the NTP had the intended effect, partly because relatively few disadvantaged pupils sampled had received tutoring, making effects hard to identify.
8. The lack of metrics appears to have been partly due to DfE’s desire to roll out the programme quickly and minimise burden. However, it meant that the NAO [could still not evaluate](#) the NTP’s value for money as late as 2023, despite its estimated whole-life cost of over £1 billion: “there is limited evidence on how extra direct funding for schools was spent and how far it was used to support disadvantaged pupils”. Any opportunity to refine the NTP to improve value for money was thus lost.
9. While we are sympathetic to the need to reduce burdens on schools, it would not have added significant burden to record the most important metric: the number of disadvantaged pupils tutored. Schools were already required to report the total pupils tutored (via the school census) so must already have recorded per-pupil data. We suggest it is more likely that this question was simply not included as part of initial programme design.

Why does this problem occur?

Metrics are recommended, not required

10. HMT publishes vast quantities of [evaluation guidance](#), in the [Green Book](#) and the [Magenta Book](#). The latter recommends that metrics and data should be considered when developing policies, saying “the collection of data required for an evaluation should be planned alongside the development of the intervention”, including “the evaluation questions to be answered [and] who can provide relevant data”.
11. However, this does not appear to be compulsory, or carried into practice on all major projects. HMT publishes a [guide to developing the Programme Business Case](#) and an [equivalent](#) for projects, but neither require project developers to define the questions to be evaluated, the metrics to be collected, or data to be used.

12. Similarly, the Infrastructure and Projects Authority publishes a [hefty list of documents](#) for assurance reviews, but this does not appear to direct that projects should be rejected for not defining metrics. Thus, business cases for major projects and programmes can be developed without specifying metrics.
13. Previous attempts have been made to tackle this issue:
 - a. An [independent report](#) by Michael Barber identified these issues with project evaluation in 2017, proposing that the Treasury make Budget allocations conditional on departments collecting evaluation data.
 - b. The Evaluation Task Force (ETF) was subsequently created with a remit to ensure evidence and evaluation sit at the heart of government spending decisions, and [reviewed](#) the evaluation plans included in departments' bids for funding as part of the 2021 Spending Review.
14. However, in a 2021 [report](#), the NAO noted that “outside of Spending Reviews, HM Treasury has few levers and exercises little scrutiny to ensure that departments comply with evaluation requirements and expectations”. The NAO warned that the main barriers to good evaluation in Government remains “a lack of incentives... together with few adverse consequences” for weak evaluation.
15. As a consequence of this, evaluation questions may not be defined upfront, and where they are, metrics and the processes required to collect them are even more rarely clearly defined.

How can it be fixed?

16. Firstly, **we recommend that business plans for major projects and programmes should be required to define the following three elements upfront:**
 - a. The main **policy questions** to be evaluated
 - b. The **metrics** to be used to answer those questions
 - c. How the **underlying data** and data systems will generate the metrics.
17. For example, if this had been done for the NTP, the programme business case would have defined:
 - a. The major questions to be evaluated, i.e. “is the NTP helping disadvantaged children receive high quality tutoring that improves outcomes”?
 - b. The major metrics to be evaluated, i.e. the number of disadvantaged children being tutored, the quality of the tutoring, and the effect on outcomes.
 - c. The data systems to record the metrics: i.e. schools could have been asked to record tutoring per pupil on the National Pupil Database from the start

(which is now done and is not burdensome). This would have made it straightforward to gather metrics on both pupil disadvantage and educational outcomes.

18. If these metrics had been available earlier, it is possible that the programme would not have been renewed in 2021 with a completely different design at a cost of hundreds of millions of pounds; or that the existing programme could have been refined and improved to more clearly deliver value for money.
19. Without this, evaluation will continue to be limited by low-quality or unavailable data, which it will be too late to remedy post hoc. However, if metrics are defined upfront, the operational elements of the programme can be designed to ensure that evaluation is low-burden.
20. We also recommend that **programme budget approval should be conditional on the above elements being in place**. This would be in line with practices in other countries that have led to clear improvements:
 - a. In Japan, the Treasury only approves spending that is backed by clear plans for impact evaluation. While this policy increased spend on evaluations, [overall expenditure fell](#) because money was targeted more effectively.
 - b. Canada introduced its [Policy on Results](#) in 2016, which integrates evaluation reporting into regular programme budgets. The Treasury Board of Canada must approve the indicators used to measure programme performance. The policy led to [government-wide improvements in evaluation](#).
21. We recommend that **analysts should be involved in the development of the project's or programme's business case**. Analytical insight is needed to understand how data and metrics can support the evaluation of the policy questions. As Professor Sir Ian Diamond, the National Statistician, recently [told](#) a Select Committee:

“The very best policy... is made by teams of policy professionals and analysts working together to identify the questions that need answers and the data that are needed to answer those questions and developing, in an iterative way, better analysis, which leads to better policy” (our emphases).
22. And finally, external oversight would be beneficial. We recommend **the NAO should be asked to comment on the evaluation plans put forward in business cases**, to ensure it has adequate data to assess value for money.

About us

23. The Centre for Public Data is a non-partisan, non-profit research organisation that works to improve the quality of data produced or regulated by the UK government. We are happy to discuss any issues raised here: contact@centreforpublicdata.org.