OFFICIAL

Australian Government
Australian Taxation Office

Is the value of evaluation hiding in the shadows?

ATO Evaluation Hub presentation to the aes26 International Evaluation Conference

17 September 2026 – Darwin | Garramilla

EXTERNAL

OFFICIAL



Acknowledgment of Country

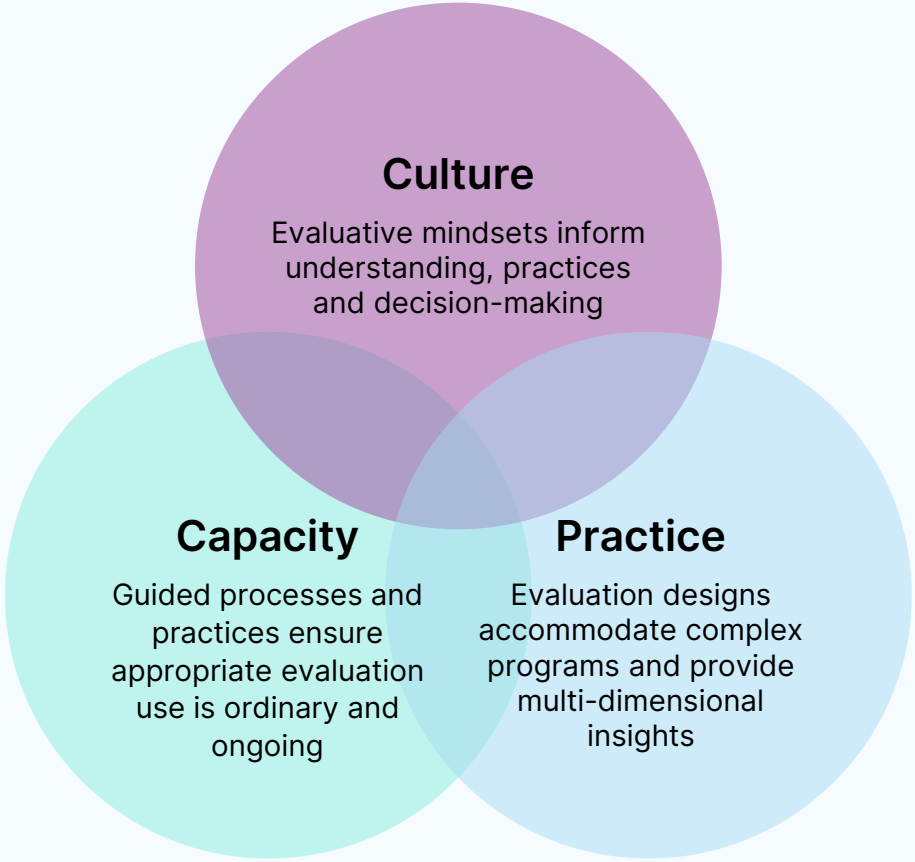
We acknowledge the Traditional Owners and Custodians of Country throughout Australia and their continuing connection to land, waters and community. We pay our respects to them, their cultures, and Elders past and present.

We recognise the unique relationship Aboriginal and Torres Strait Islander peoples have to Country, culture, community and the important role this plays in us all walking together as Australians.

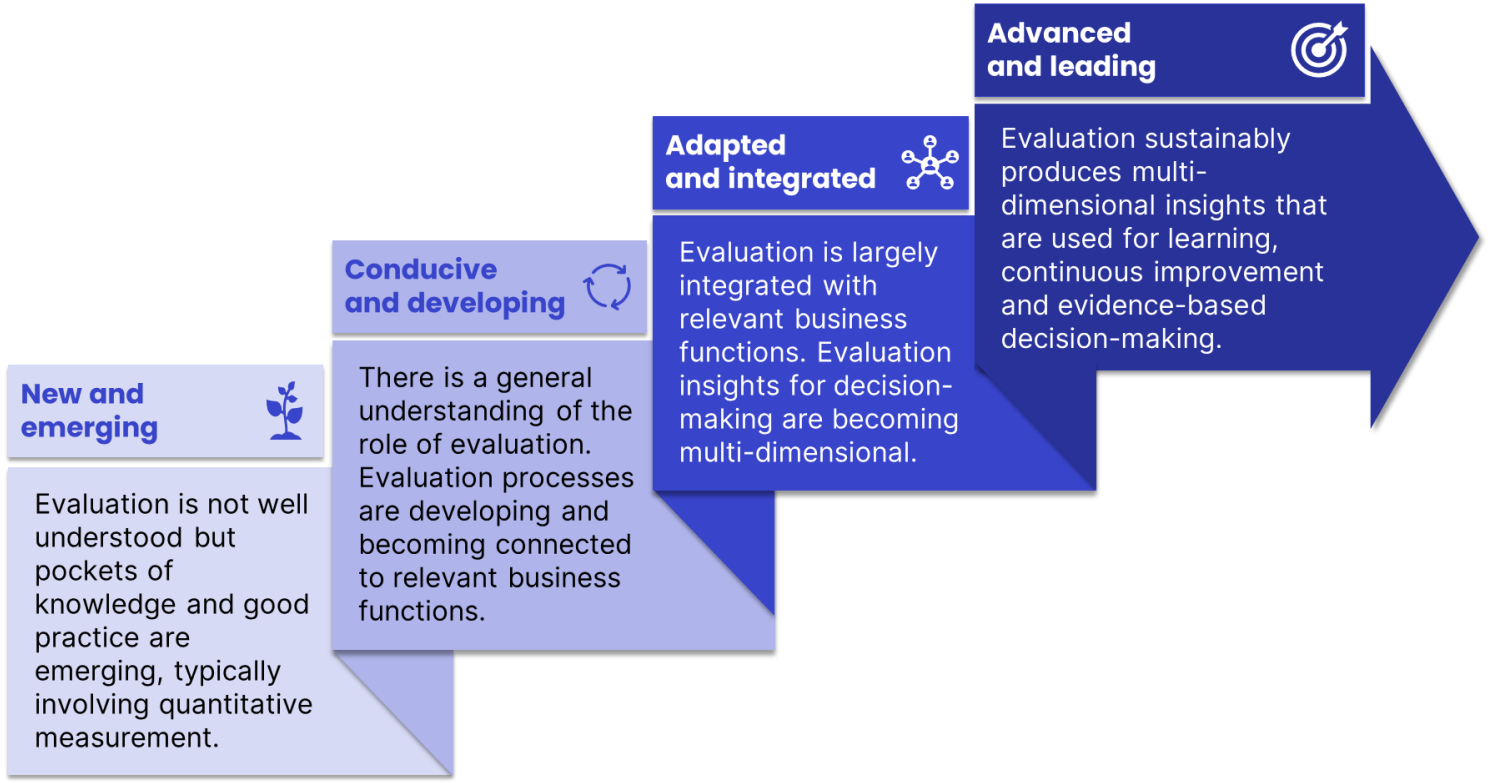
We value the contribution Aboriginal and Torres Strait Islander peoples make to our organisation and the broader tax, superannuation and registry systems.

The role of the ATO Evaluation Hub

The Evaluation Hub is a driver for building ATO evaluation maturity …



Dimensions of evaluation maturity



… from a new and emerging level to an aspirational state of advanced and leading.

Our hub and spoke operating model connects us to related functions across the ATO …



- Leading the ATO Evaluation Community of Practice
- Building evaluation capability
- Advising on fit-for-purpose evaluation
- Connecting with internal and external evaluation networks

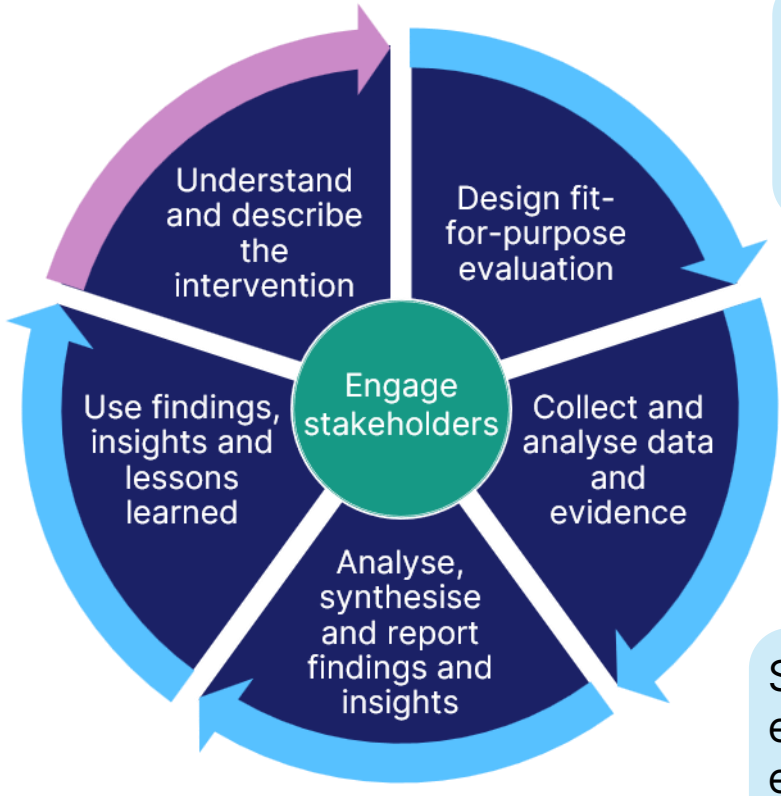
… and our services connect us with ATO people and business areas and the wider evaluation profession.

Evaluative thinking and techniques are embedded across the evaluation life cycle

Logic modelling to uncover, critically assess and describe how an intervention (policy/program/project) operates to bring about change

Using evaluation findings and insights to inform ongoing decision making and to support evidence base for problem/ situation/environment analysis

Tailoring communications to facilitate learning and translation of findings and insights into action



Evaluation planning as a creative, inclusive and reflective exercise to make choices about what, when and how to evaluate

Selecting and analysing data and evidence that are relevant for monitoring progress and making evaluative judgments

Synthesising and interpreting evidence to form judgments against evaluation questions or criteria and about the achievement of outcomes

OFFICIAL

The advantages of this approach



Creates value and usable insights throughout the life cycle of a program or intervention



Reflects the reality that not all programs will be evaluated



Builds evaluative thinking and technical skills in program leads and other stakeholders

OFFICIAL

Where value is created throughout the program life cycle

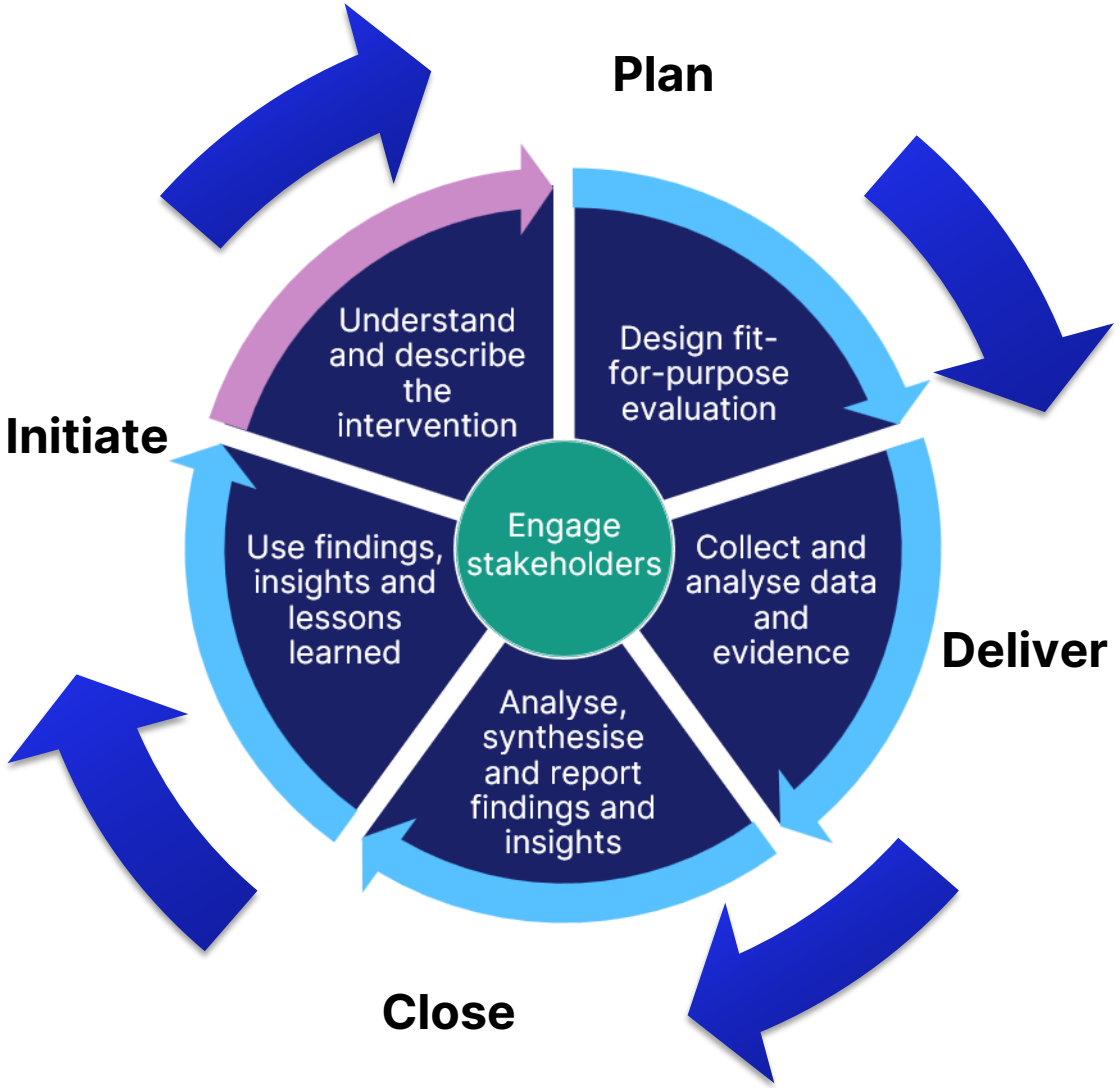
Evaluation evidence:

- informs direction setting and adaptation to changing environments
- contributes to organisational knowledge and evidence base for analysing problems and exploring options
- informs resource allocation and governance requirements

Evaluation reveals what works, why and for whom and confirms or challenges program logic

Sharing findings widely:

- facilitates learning and translation into action
- highlights achievements and improvement opportunities
- demonstrates transparency and accountability



Logic modelling:

- aids prioritisation and desktop 'road testing' of options
- builds evaluative thinking skills and understanding among stakeholders
- fosters realistic expectations, outcomes and timeframes
- produces artefacts for validation and communication
- provides foundation for good evaluation design and planning

Good evaluation design enables relevant, credible and useful findings and insights

Monitoring flags any need for adjustment early and assures stakeholders that progress and outcomes are being tracked

We found Gullickson's discussion of defining evaluation to 'fully describe, fully judge' resonated with us and our approach (2020, p.2)

The 'fully describe, fully judge' definition draws on Stake (1977, p.374) to outline that:

'...the task of evaluation is to provide a full description of the evaluand, a justified set of criteria and standards by which to judge it, and a process that enables generation of a judgement.'

'To flesh it out, I have expanded Scriven's original logic (Gullickson, 2018) based on Nunns (2016) to make the defining, judging, and the inherent reasoning aspects more visible...'

(Gullickson 2020, pp.2)

OFFICIAL

This mapping to the evaluation process is also a helpful way to name and demonstrate to our stakeholders the value created from evaluation

As Gullickson (2020, p.4) states:

'...we can explain to clients how what we have done, if it is more on the "fully describe" side, fits into the general process of evaluation.'

Stakeholders often perceive value is only delivered here, but all steps are valuable

The Expanded Logic of Evaluation - extract from Gullickson (2020, Table 2)

Step	Mapping
0. Clarify evaluation purpose and assess evaluability	
1. Define the evaluand	Fully describe
2. Define the group to which the evaluand belong	Fully describe
3. Identify criteria/ delineate evaluation questions	Fully judge
4. Identify performance standards	Fully judge
5. Justify the criteria and standards	Fully judge
6. Measure; observe evaluand's performance	Fully describe and fully judge
7. Justify the measures	Fully describe and fully judge
8. Synthesise an evaluative judgement	Fully judge
9. Justify the synthesis method	Fully judge
10. Report judgement	

EXTERNAL

OFFICIAL

Our approach aligns with our aim for evaluation thinking and techniques to be applied upfront in program design and recognises that not all programs will be or need to be evaluated

We need to prioritise limited evaluation capacity and capability to where it is most valuable.

As the Commonwealth Evaluation Policy states:

'The choice to evaluate and the scale of effort and resources allocated to an evaluation should be proportional to the value, impact, strategic importance and risk profile of the program or activity.'

(The Treasury n.d.)

It also reflects our vision to build evaluation capability and a learning culture across the organisation

We are building thinking skills and a foundation for program leads and other stakeholders [our clients] to make evaluative judgments later in the evaluation process.

Gullickson (2020, p4) talks about using that 'specialised knowledge evaluators would have under this [fully describe, fully judge] definition…to explicitly prepare them [clients] to do it on their own'.

We see this as a sustainable benefit where clients will be thinking differently for future investment decisions and future program designs – and to design fit-for-purpose evaluation where warranted.

OFFICIAL

How do we demonstrate what that looks like in action?

A case study from the Shadow Economy Program

OFFICIAL

The Shadow Economy Program

- In operation since 2018-19 in various forms and extended through several Budgets
- Funds a range of ATO activities intended to address aspects of the shadow economy

But...

- **Shadow economy behaviour is evolving influenced by economic, social and technological changes**
- **Government expectations to evaluate and demonstrate impact are increasing**

High-level logic of the Shadow Economy Program

Unpack the problem to be addressed (related to the **drivers** of shadow economy behaviour)

Connect activities invested in by the program to the problem and outcomes

Define the outcomes to reflect the **positive changes** that the program is seeking to bring about to address the problem



Assumptions

Identify assumptions about how the program should logically work and critically **assess** whether they are sound

Defining the drivers of shadow economy behaviour



Poor transparency of shadow economy behaviour makes detection and identification difficult



Risk-benefit (chance of getting caught and severity of consequences vs benefit) favours shadow economy participation



Barriers to complying with the tax and super law make it easier to participate in the shadow economy than not



Perceived bias that others are unfairly gaining from the system



Acceptance of shadow economy behaviour as a **social norm** in certain industries and communities

Connecting outcomes across the Shadow Economy Program

This started with redefining the program outcomes so that they:

- explicitly connect addressing drivers of shadow economy behaviour to the intended outcomes of the program, and to the ATO's and Government's broader goals, through a logical causal chain
- help to focus activities on what we can control and on changes we can bring about
- provide the program management office and business areas with a practical way to manage the program and the activities the program invests in
- provide a sound foundation for evaluation

Drivers of shadow economy behaviour	**Behaviour description**	**Investment level outcomes** (single strategy lens)	**Whole of program level outcomes** (combined efforts across shadow economy strategies)	**Broader goals** (ATO contributes to but does not control)
1. Poor transparency of shadow economy behaviours makes detection and identification difficult	The hidden nature of shadow economy activities, such as cash transactions or off-the-books employment, makes them difficult to detect, track and quantify. This limits the effectiveness of enforcement and policy responses.	**IO1** Use data and digital tools to detect where people are engaging with shadow economy behaviour and identify participants.	**PO1** The ATO uses intelligence on shadow economy behaviour to coordinate cross agency approaches that address shadow economy behaviour.	Shadow economy behaviours are not hidden from regulators. Businesses are transparent about their positive compliance with their obligations to the community and other businesses.

Using the program logic to inform program investment decisions

- Decisions for future program investment are informed by:
  - the alignment of current activities to the program outcomes and identified gaps for possible further investment
  - the need to ensure a balanced portfolio of investments that maximise the effectiveness and efficiency of the program.
- Proposals for new program activities need to clearly explain how they will address one or more behavioural drivers and how they will logically contribute to the program outcomes.
- This process provide a sound foundation for planning fit-for-purpose monitoring and evaluation and for further research to:
  - demonstrate how well the program is addressing shadow economy behaviour and identify what's working or not and why
  - understand community attitudes and perceptions of shadow economy behaviour and actions the ATO takes to address it.

Assessment criteria

The proposal clearly demonstrates the:

1. problem is linked to the shadow economy behaviour/ drivers and is supported by relevant evidence
2. outcomes are aligned to addressing the problem and contribute to the Program outcomes
3. actions are logical to achieve the proposal outcomes and can be delivered, and
4. progress and performance of the proposal will be monitored to inform changes as needed, and evaluated to inform evidence-based decisions

What happened when a business area tested the logic of some existing activities under the Shadow Economy Program?

- When a business area evaluation team worked through logic modelling with owners of some activities running under the Shadow Economy Program, it became clear that those activities didn't connect well with their stated outcomes nor with addressing drivers of shadow economy behaviour.
- They had to call it out – there wasn't a solid logical foundation for the activities nor for planning evaluation.
- They went back to basics to redesign the activities – asking:
  - What problem or risks are the activities seeking to address?
  - What actions are being taken and why?
  - Where are the gaps to achieving the stated outcomes?
  - How do the activities contribute to the Shadow Economy Program outcomes and ATO strategic objectives?
  - How will we know if the activities are effective?

This practical experience helped the business area understand the value of using evaluation thinking and tools upfront…

…and the need to build evaluation capability and culture.

- The activity owners and evaluation team had recognised, and could explain to others, that without the upfront use of evaluation thinking and tools at the design stage that the business area risked:
  - investing in activities without knowing whether they're likely to work or are working
  - relying on measures of outputs to judge performance rather than evidence of outcomes
  - being unable to confidently demonstrate the value and impact of their work.
- The evaluation team reflected on how their experience, and the broader need, is about more than building evaluation capability – it's also about culture and shifting from a mindset of delivering activity to delivering impact. That shift would strengthen program activities, decision-making and ultimately outcomes for the community.

OFFICIAL

While this case study is a positive one in the program context, the challenge of demonstrating the full value of evaluation remains more broadly.

We confronted this challenge in a recent evaluation of our own services.

Our evaluation followed a theory-based approach that linked available evidence to a theory of change that asserts a logical causal connection between our activities and a growth in evaluation culture, capacity and practice.

Our overall question – **How had the Evaluation Hub contributed to building evaluation maturity?** – sought to clearly align it with our theory of change and away from the idea that our value can and should be measured by the number of completed evaluations we facilitate.

EXTERNAL

OFFICIAL

How does the Evaluation Hub contribute to building evaluation maturity?

What we can influence and contribute to

The extent to which evaluation maturity is built will be influenced by the:

- *organisation's learning culture and capacity*
- *whole of APS evaluation reform priorities*

What we can control



Broader goals

Intermediate outcomes
Building evaluation maturity

Immediate outcomes
Meeting evaluation and learning needs

Theory informed activities

Theory informed foundations

ATO is a leading tax, superannuation and registry administration known for our contemporary service, expertise and integrity (ATO Corporate Plan 2024-25)*

The ATO has the capability to do its job well (APS Reform Agenda)

Culture
Evaluative mindsets inform understanding, practices and decision-making in the ATO

Capacity
Guided processes and practices ensure evaluation and its appropriate uses are ordinary and ongoing

Practice
Evaluation designs are able to accommodate complex programs and provide multi-dimensional insights where needed

Our guidance and support helps business areas to identify and meet their needs for evidence based decision making through fit-for-purpose evaluation practices

Our evaluation capability building activities are relevant and interesting, and help to meet individual and business area learning needs

Our leadership of the ATO Evaluation Community of Practice and engagement connects and shares learning between evaluation support areas

Expert guidance and support

Evaluation capability building (ECB)

Community of Practice

Internal and external engagement

Service delivery strategies
Modelling and coaching, two-way learning, client led

Service principles
Collaboration, utilisation, fitness for purpose, learning, evaluative thinking

** The ATO vision was updated in March 2025 after development of the logic model.*

We approached this bigger question through more specific questions

Key evaluation questions

- How well do our advice and support services meet business areas' needs?
- How well do Evaluation Hub workshops meet learners' needs?
- How relevant and interesting are ATO Evaluation Community of Practice (COP) events for participants?
- What signs of growth in evaluation maturity are evident from Evaluation Hub activities?

Data and evidence sources

- Feedback, observations, review of documents and emails relating to evaluation advice and support
- Case studies of intensive evaluation advice and support and the difference it made
- Attendance data, feedback survey responses, document review and observations relating to COP events and our workshops
- Responses to follow-up surveys and interviews of workshop participants on the application of their learning

The findings support an overall judgment that together our activities have contributed to building ATO evaluation maturity

- **Our guidance and support has helped business areas identify and meet their evaluation and business needs**
  - formal and unprompted feedback offered during and after engagements indicates that our services are appropriately tailored toward meeting business area needs
  - our intensive support services made observable differences connecting evaluation with relevant business processes, building capabilities, and informing decision-making
- **Our capability building activities are relevant and interesting and help to meet individual and business area learning needs**
  - there's a high level of participation in our workshops, indicating relevance to peoples' interest and learning needs – supported by formal feedback provided by some participants following workshops
  - most people who told us they'd put the logic modelling workshops concepts into practice said it made tangible differences to informing and improving business practices and to their own decision-making
- **People are connecting and sharing learning through the ATO Evaluation COP**
  - COP participation continues to grow indicating it's perceived as relevant to participants' interests or roles – which is supported by formal feedback from participants following COP meetings
- **We also identified areas for further reflection and action to add more value for learners and internal clients and for improving evaluation practice and culture**

OFFICIAL

Closing thoughts

If ‘fully judging’ is what distinguishes evaluation from other forms of structured inquiry – what happens when we claim value derived from ‘fully describing’ as being value derived from evaluation?

OFFICIAL

OFFICIAL

As McDonald, Rogers and Kefford (2003) recommend, it's important to both allow sufficient time to build evaluation capacity and 'to quickly and repeatedly get "points on the board" to be seen as … of some use immediately'.

OFFICIAL

Thank you

Contact us at evaluation@ato.gov.au

References

Gullickson, A. (2020). The whole elephant: Defining evaluation. Evaluation and Program Planning, 79. https://doi.org/10.1016/j.evalprogplan.2020.101787

McDonald, B., Rogers, P., & Kefford, B. (2003). Teaching People to Fish? Building the Evaluation Capability of Public Sector Organizations. Evaluation, 9(1), 9-29. https://doi.org/10.1177/1356389003009001002

The Treasury (n.d.). Commonwealth Evaluation Policy. Australian Centre for Evaluation website, accessed 14 September 2026