



Social Investment Agency
Toi Hau Tāngata

Guidance for the Impact Measurement Standards

June 2026



New Zealand Government
Te Kāwanatanga o Aotearoa



Authors

Dr Craig Jones, Dr Sarah Ann Pinto, Nataly Noguera Blue, Dr Lisa Houghton, Dr Zachary Penman and Dr Michelle Barr

Acknowledgements

Special thanks to the agencies and people who helped us in the consultation process. We are grateful for the technical working group and many others who contributed to the development of these standards: Charlie Russell, Claudia Recker, Darin Bishop, Gabrielle Po Ching, Professor Gail Pacheco, Gareth Edwards, Gemma Greentree, Hamed Nazari, Dr Jas Qadir, Kirsten Jensen, Megan Davis, Dr Meghan Stephens, Megan Somerville, Michelle Bly, Oliver Parsons, Peter Dolan, Dr Phillip Hall, Dr Robyn Whittaker, Dr Rosie Dobson, Sally Page, Sarah Crichton, Sankar Ramasamy, Dr Seán Barnes, Siamak Dehghan, Dr Steve Murray, Theresa Lealofi and Professor Tracey McIntosh. We appreciate there are a multitude of views on how to position these standards. Final judgement calls and errors of fact or interpretation rest with the Social Investment Agency.

Creative Commons Licence



This work is licensed under the Creative Commons Attribution 4.0 International licence. In essence, you are free to copy, distribute and adapt the work, as long as you attribute the work to the Crown and abide by the other licence terms. Use the wording 'Social Investment Agency' in your attribution, not the Social Investment Agency logo.

To view a copy of this licence, visit creativecommons.org/licenses/by/4.0.

Liability

While all care and diligence has been used in processing, analysing and extracting data and information in this publication, the Social Investment Agency gives no warranty it is error free and will not be liable for any loss or damage suffered by the use directly, or indirectly, of the information in this publication.

Citation

Social Investment Agency 2026. Impact Measurement Standards Guidance. Wellington, New Zealand.

ISBN 978-1-997329-04-6 (online)

Published in June 2026 by
Social Investment Agency
Wellington, New Zealand

Contents

Purpose of the Impact Measurement Standards	4
Applying the Impact Measurement Standards	5
Who do the standards apply to?.....	5
How should you use the standards?	5
What do the key concepts mean?.....	6
Measuring the impact of programmes and services	7
What are the three levels of impact evidence?	7
Expected practice across the standards	13
Measuring value.....	15

Purpose of the Impact Measurement Standards

The Social Investment Agency is mandated to set standards for social investment. The Social Investment Agency has developed the Impact Measurement Standards to help government agencies make better decisions about public investment in social outcomes.

The standards provide a shared framework for assessing:

- what programmes and services are trying to achieve
- whether they are making a meaningful difference
- for whom outcomes are improving.

Used well, the standards support more consistent, credible, and transparent use of evidence across government.

Improving outcomes for New Zealanders

Better evidence helps agencies understand what works, for whom, and in what circumstances.

This supports:

- continuous improvement in services
- better targeting of resources
- stronger long-term outcomes
- reduced inequities over time.

Supporting accountability

The [Public Finance Act 1989](#) requires agencies to ensure public resources are used effectively and efficiently, and to advise on the performance of public spending.

As a Te Tiriti o Waitangi partner, the Crown also has responsibilities to improve equitable outcomes for Māori.

These standards support agencies to meet these responsibilities through stronger, more consistent use of evidence on outcomes and impact.

Purpose of this document

This document provides practical guidance for government agencies on how to apply the standards when making commissioning, investment, evaluation, and service improvement decisions.

This version of the Impact Measurement Standards is an updated release. However, the standards will continue to evolve based on practical learning from use across government and Social Investment Fund Initiatives.

Applying the Impact Measurement Standards

Who do the standards apply to?

These standards are initially intended for government agencies that fund, commission, deliver, or oversee programmes and services designed to improve social outcomes. They may also be useful to delivery partners, evaluators, and other organisations involved in planning, measuring, or reviewing outcomes.

Government agencies

The standards are relevant across agencies working in areas such as child wellbeing and protection, education, health, housing, justice, policing, employment, disability, and social development.

Within agencies, they are most relevant to teams responsible for:

- commissioning and procurement
- policy and strategy
- finance and investment decisions
- research, insights, and evaluation
- monitoring and performance
- operational leadership.

Applying the standards well often requires collaboration across these functions.

How should you use the standards?

Agencies should use the standards to support better decisions across the lifecycle of programmes and services, including to:

- assess the strength of existing evidence
- identify gaps in evidence and prioritise improvement
- set proportionate evidence expectations for new investments
- inform funding, continuation, redesign, or scaling decisions
- compare options alongside strategic priorities and practical considerations
- strengthen accountability for outcomes achieved.

The standards should be applied pragmatically. Evidence expectations should be proportionate to the size, risk, maturity, and strategic importance of the investment.

The Social Investment Agency will use the standards to support agencies applying a social investment approach and to strengthen evidence expectations in funding and commissioning processes.

What do the key concepts mean?

Term	Definition
Impact evaluation	Impact evaluation assesses whether a programme or service has made a meaningful difference and examines the extent to which those changes can be credibly linked to the intervention through appropriate methods.
Impacts	Impacts are the broader or longer-term changes associated with a programme, service, or policy. These may be social, economic, cultural, environmental, or system-level, can be positive and negative, intended or unintended.
Programme	A programme refers to a set of activities, services, or interventions designed to achieve one or more outcomes. A programme may be delivered locally, regionally, or nationally, and may range from a discrete initiative to a broader strategy or service model (for example, Kia Manawanui Aotearoa: Long-term pathway to mental wellbeing).
Outcomes	Outcomes are the changes experienced by individuals, families, whānau, or communities during or after a programme or service. These are typically short - to medium-term changes in people’s lives, behaviours, capabilities, circumstances, or wellbeing.
Causal attribution	Causal attribution refers to the extent which observed changes can be linked to a programme, service, or policy rather than to other factors. It helps decision-makers understand whether outcomes were caused by the intervention, in whole or in part. Read Overview: Strategies for causal attribution for more information about causal attribution.
Contribution analysis	Contribution analysis is an approach used to assess whether, and how, a programme plausibly contributed to observed outcomes, particularly when experimental or counterfactual methods are not feasible.
Kaupapa Māori evaluation	Kaupapa Māori evaluation is an approach to evaluation grounded in Māori values, mātauranga Māori, and Te Ao Māori worldview. It centres Māori voices, priorities, and ways of knowing throughout the evaluation process. It is particularly important when evaluating programmes, services, or outcomes involving Māori communities. ¹
Process evaluation	Process evaluation examines how a programme, service or policy was implemented, including what was delivered, how it operated in practice, and the factors that supported or constrained successful delivery.
Theory of Change	A Theory of Change explains how and why a programme, service, or policy is expected to lead to desired outcomes or impacts in a particular context. It sets out the pathway from activities to short, medium, and long-term change, including the key assumptions that underpin success. A Theory of Change may be summarised through a logic model, diagram, or other visual framework.
Social investment	Social investment is an approach to improving social outcomes by using evidence to understand what works for whom, where, at what cost, and for what benefit, and then using that evidence to guide investment, commissioning, and service improvement decisions.

¹ Cram, F., Pipi, K., & Paipa, K. (2018). Kaupapa Māori evaluation in Aotearoa New Zealand. In F. Cram, K. A. Tibbetts, & J. LaFrance (Eds.), *Indigenous Evaluation*. *New Directions for Evaluation*, 159, 63–77.

Measuring the impact of programmes and services

Impact measurement helps agencies understand whether programmes and services are making a meaningful difference, for whom, and with what level of confidence.

The standards describe three levels of impact evidence. These levels reflect progressively stronger evidence about outcomes and causal impact, while recognising that the most appropriate level depends on the programme's maturity, scale, risk, and decision context.

- Level 1 focuses on a credible Theory of Change grounded in evidence.
- Level 2 focuses on measuring observed changes in outcomes or impact.
- Level 3 focuses on using robust methods to assess whether the programme contributed to or caused those changes.

All programmes and services are expected to:

- meet Level 1 standards
- seek Level 2 evidence once implemented, where practical and proportionate
- consider whether Level 3 is feasible and worthwhile.

There is also expected practice across all levels ([see Expected practice section](#))

What are the three levels of impact evidence?

Level 1: A Theory of Change based on evidence

All programmes and services are expected to meet Level 1 standards.

Level 1 focuses on having a clear and credible Theory of Change that explains how a programme or service is expected to achieve desired outcomes or impact, and the evidence and assumptions that underpin this logic.

Level 1 is particularly important for new initiatives, but it is also valuable for existing programmes where delivery models, target populations, or intended outcomes need to be clarified or reviewed.

A strong Level 1 foundation improves programme design, commissioning, performance measurement, and future evaluation.

To meet Level 1 standards, agencies will need to:

1. Define the pathway to change

Clearly describe how programme inputs, activities, and outputs are expected to lead to short-medium-, and long-term outcomes and impacts.

2. Use evidence and stakeholder insight

Draw on available local and international evidence, operational experience, and input from partners, communities, and stakeholders to inform the Theory of Change. Be clear about the strength, relevance, and limitations of the evidence supporting the proposed pathway to change. Use Kaupapa Māori approaches where appropriate.

3. State assumptions and risks

Document key assumptions, contextual factors, risks, and dependencies that may influence whether intended outcomes are achieved.

4. Identify how success will be measured

Specify the priority outcomes or impacts to be measured, and how evidence will be collected over time.

5. Consider equity and differential outcomes

Consider whether outcomes may vary for different groups such as Māori, Pacific peoples, disabled people, gender groups, geography, income, or other relevant populations.

Useful resources

- [Understanding Theory of Change and logic models](#)
- [Evaluation Handbook Dec 2017.pdf \(dpmc.govt.nz\)](#), pp. 15-25
- <https://evaluationhub.education.govt.nz/theory-of-change/a-theory-of-change/>
- <https://whatworks.org.nz/logic-model/>
- The [Social Investment Fund Quick Guide: Understanding how we use outcomes](#) the [Living Standards Framework](#) and [He Ara Waiora](#) are useful starting points to help identify outcomes.

Example: E Tū Whānau Mahere Rautaki: [Framework for Change](#) is a good example of Level 1 evidence that is strengths-based and guided by the Te Ao Māori worldview.

Level 2: Evidence indicates changes in outcomes or impacts

Programmes and services that are established or operating in New Zealand should generally seek Level 2 evidence where practical and proportionate.

Level 2 focuses on measuring observed in outcomes and impacts for those receiving a programme or service, using credible data linked to the Theory of Change.

Level 2 evidence can provide strong insight into whether outcomes are improving, for whom, and over time. While it may not fully isolate the effect of the programme from other influences, it is often highly valuable for performance improvement, accountability, and commissioning decisions.

To meet Level 2 standards, agencies will need to:

1. Measure priority outcomes

Collect and report evidence on short-, medium-, and where feasible longer-term outcomes or impacts linked to the Theory of Change.

2. Use quality data

Use measures that are valid, reliable, timely, and fit for purpose, with attention to completeness and consistency.

3. Measure at an appropriate level

Capture information at individual level, whānau, household, community, or other suitable levels depending on the nature of the programme and what is ethical and practical.

4. Examine differences across groups

Assess whether outcomes vary across relevant population groups such as Māori, Pacific peoples, disabled people, gender groups, geography, income, or other priority populations.

5. Explain limitations and context

Clearly state important caveats, data limitations, contextual factors, and other influences that may affect interpretation of results.

6. Use findings for improvement

Use evidence to support service improvement, commissioning discussions, and future evaluation planning.

Useful resources

- <https://www.procurement.govt.nz/guides/social-services-procurement/developing-a-social-services-procurement-plan/how-to-measure-outcomes-and-outputs/>
- [SIF-Quick-Guide-Continuous-Learning](#)
- [Evidence and evaluation | Department of the Prime Minister and Cabinet \(DPMC\)](#)

Example: The Ministry of Health's [Mana Ake – Stronger for tomorrow](#) report is a good example of Level 2 evidence. It also offers a useful insight into the purpose of evaluation to guide future decisions.

Level 3: Evidence provides high confidence causal impact

Level 3 focuses on using robust methods to provide a high degree of confidence about whether a programme or service caused, or materially contributed to, observed outcomes or impacts.

This level is most appropriate where decisions are high value, high risk, strategically significant, or where compelling evidence is needed to support scaling, redesign, or major reinvestment decisions.

Level 3 evidence is not required for every programme. Agencies should consider whether it is feasible, proportionate, ethical, and useful for decision making.

A range of methods may be appropriate, including experimental, quasi-experimental, theory-based, and mixed-method approaches.

To meet Level 3 standards, agencies will need to:

1. Use robust causal methods

Use methods capable of credibly assessing whether observed outcomes were caused by, or materially linked to, the programme rather than other factors.

2. Consider alternative explanations

Identify and test other plausible factors that may explain observed changes.

3. Consider suitable comparison logic, where relevant

Where appropriate, use counterfactuals, comparison groups, staggered rollout, matched groups, contribution analysis, or other robust approaches suited to the intervention and context.

4. Use quality data

Use reliable and fit-for-purpose data at an individual, whānau, community, or other appropriate level.

5. Address bias

Ensure that analysis appropriately identifies, manages, and tests potential sources of bias.

6. Examine differences across groups

Consider whether impacts differ across relevant population groups such as Māori, Pacific peoples, disabled people, gender groups, geography, income, or other priority populations.

7. Be transparent and replicable

Clearly document methodology, assumptions, limitations, and sources of uncertainty so findings can be understood and scrutinised.

Level 3 is easier to achieve when:

- The **programme is new, expanding, or being redesigned**, allowing evaluation approaches to be built in from the start.
- The **programme is sufficiently large in scale or reach**, so meaningful comparisons and robust analysis are possible.
- **Demand exceeds available places or rollout can occur over time**, creating fair opportunities for phased implementation or comparison approaches.
- **Relevant outcome data already exists, or can be collected reliably**, for both participants and suitable comparison groups where needed.
- **Eligibility criteria, referral pathways, or access processes are clear**, making it easier to identify comparable groups or test causal pathways.
- **The expected benefits of stronger evidence justify the investment**, particularly where significant funding or major decisions are involved.
- **Ethical and operational considerations allow stronger evaluation methods** without undermining service delivery or fairness.

Prioritising programmes and services for Level 3 impact measurement

Level 3 impact measurement can require significant time, capability, and resources. It should therefore be prioritised where stronger causal evidence is likely to add the greatest decision-making value.

When deciding where to invest in Level 3 evidence first, agencies should consider the:

- **Scale of investment** - prioritise programmes or services involving substantial or ongoing public investment.
- **Significance of upcoming decisions** – focus on areas where major decisions are approaching, such as continuation, expansion, redesign, reprioritisation, or contract renewal.
- **Strength of existing evidence** – prioritise programmes where current evidence is weak, uncertain, contested, or outdated.
- **Feasibility and evaluability** – focus where robust methods are practical, ethical, and supported by suitable data.
- **Potential wider learning value** - prioritise evaluations where findings could inform multiple programmes, sectors, or future investment decisions.
- **Equity and strategic importance** – consider areas with significant implications for Māori outcomes, reducing inequities, or achieving wider government priorities.

Investing in impact measurement

Investment in impact measurement should be proportionate to the scale, risk, complexity, and strategic importance of the programme or service.

Stronger evidence can improve decision-making, but measurement activity also has costs. Agencies should balance the expected value of additional evidence against the time, resources, and operational burden required to generate it.

In practice:

Higher investment in impact measurement may be justified when:

- public investment is substantial or ongoing
- significant continuation, scaling, or redesign decisions are pending
- uncertainty about effectiveness is high
- findings could inform wider policy or system decisions
- risks of ineffective spending are material.

Lighter-touch approaches may be appropriate when:

- programmes are small-scale or time-limited
- risks are lower
- existing evidence is already strong
- the cost of complex evaluation would outweigh likely learning value.

Planning and costing

Agencies should plan evaluation requirements early and budget for them from the outset. An activity-based costing approach can help align evaluation effort with programme scale and evidence needs.

Useful resources

- <https://thehub.sia.govt.nz/resources/making-sense-of-evaluation-a-handbook-for-everyone/>
- <https://thehub.sia.govt.nz/resources/making-sense-of-evidence-a-guide-to-using-evidence-in-policy/>

Example: [Kā Au Kahuraki Tamai Safe Homes and Communities](#) is a good example of strong Level 2 evidence, with emerging Level 3 elements. It demonstrates an evidence-based theory of change, measures outcomes before and after implementation, and uses comparison groups to understand its contribution to observed impacts.

Expected practice across the standards

The standards should be applied in the context of good commissioning, evaluation, and public decision-making practice.

Strong impact measurement depends not only on technical methods, but also on meaningful engagement, appropriate use of different knowledge systems, and balanced use of evidence in decision-making.

Across all levels of standards, agencies should seek to:

- engage communities, partners, and stakeholders appropriately
- use evidence approaches that are inclusive and fit for purpose
- recognise the value of Kaupapa Māori and other knowledge systems
- combine quantitative, qualitative, operational, and lived-experience evidence
- use findings to improve services and future investment decisions
- apply the standards proportionately and pragmatically.

Engage communities and stakeholders

Impact measurement is strengthened by meaningful engagement with communities, service users, delivery partners, and other stakeholders.

Engagement helps agencies understand which outcomes matter most, how services are experienced in practice, what barriers exist, and how evidence can be used to improve results.

Depending on the context, relevant stakeholders may include individuals, whānau, iwi and Māori partners, community organisations, delivery providers, and other public agencies.

Agencies should seek to:

- engage early when designing programmes, outcomes frameworks, or measurement approaches
- understand community priorities and definitions of success
- involve stakeholders in interpreting findings and identifying improvements
- use culturally appropriate and accessible engagement methods
- create feedback loops so evidence leads to visible learning and change.

Engagement should be genuine, proportionate, and fit for purpose.

We recommend you consider The Policy Project's [Community engagement](#) resources and Te Arawhiti's [Guidelines for engagement with Māori](#) when planning engagement.

Be inclusive of different knowledge systems

The standards are method-neutral and recognise that credible evidence can be generated through different knowledge systems, methodologies, and ways of understanding change.

This includes quantitative, mixed-method, and culturally grounded approaches such as Kaupapa Māori evaluation and evidence informed by mātauranga Māori.

The most appropriate approach will depend on the programme, population, outcomes of interest, and decision context.

For programmes involving Māori communities or outcomes, agencies should ensure evidence approaches appropriately reflect Te Tiriti o Waitangi responsibilities and enable Māori perspectives, priorities, and definitions of success to be meaningfully represented.

Agencies should seek to:

- choose methods that are fit for purpose and context
- value lived experience and community knowledge alongside administrative or statistical evidence
- involve Māori appropriately in the design, interpretation, and use of evidence where relevant
- recognise that no single method captures all dimensions of value or impact.

Inclusive evidence practice strengthens both legitimacy and decision quality.

Kaupapa Māori evaluation is about doing evaluation a Māori way, which is especially important for programmes delivered by Māori for Māori. Kaupapa Māori evaluations often use methods from western evaluation practice, however, the 'methodology and the paradigm within which Kaupapa Māori evaluators operate are Māori'.¹ Kaupapa Māori evaluators also work with organisations to incorporate traditional decision-making ways into how they capture outcomes.²

Kaupapa Māori evaluation is often less about judgement and more about finding out how and why organisations are working for the benefit of their clients, whānau and communities.³ It is expected that these different types of evaluation (like process evaluation – see below section) will be integrated with quantitative evidence of impact in decision making.

The integration of different knowledge systems is important considering partnership obligations between Māori and kawanatanga under Te Tiriti o Waitangi. There are numerous books and resource guides to support bicultural approaches to knowledge development, including Professor Angus MacFarlane's seminal work, *He Awa Whiria*.⁴ Superu's [Bridging Cultural Perspectives](#) is also a useful resource.

Integrate different types of evidence in decision-making

Impact measurement should inform decisions but should not be used in isolation.

Good public decisions are strengthened when agencies consider multiple sources of evidence together, recognising that different forms of evidence answer different questions, for example:

² Cram, F., Pipi, K., & Paipa, K. (2018). Kaupapa Māori evaluation in Aotearoa New Zealand. In F.Cram, K. A.Tibbetts, & J.LaFrance (Eds.), *Indigenous Evaluation. New Directions for Evaluation*, 159, 63–77. ²

³ [Katoa Ltd - Kaupapa Māori Evaluation](#)

⁴ See, for example, an excellent recent collection of practical examples of this model published by MacFarlane, A., Derby, M. & MacFarlane, S. (2024). *He Awa Whiria: Braiding the knowledge streams in research, policy and practice*. Canterbury University Press.

- **Impact evidence** helps show whether a programme or service is making a meaningful difference.
- **Value evidence** helps assess whether benefits justify the investment.
- **Process evaluation** helps explain how implementation influenced the results.
- **Operational and performance data** helps identify service trends and delivery issues.
- **Mātauranga Māori, whānau voice, and community** insights help ensure decisions reflect lived experience, cultural context, and what matters to communities.
- **Strategic, fiscal, policy considerations** help determine wider priorities and feasibility.

Agencies should weigh these sources together when deciding whether to continue, improve, expand, redesign, or stop programmes or services.

No single number, method, or evidence source should determine decisions on its own. Balanced judgement remains essential.

Measuring value

Value measurement quantifies the benefits of an investment, capturing both monetary and non-monetary returns. Depending on the objective, value may be assessed through measures of effectiveness, efficiency, equity and economy, or a combination of these.

These measures enable evidence-informed judgements about whether an investment aligns with the strategic objectives of the investor and support comparisons across different investments. Value measurement can be done before or after an investment is made to support decisions at different points in policy, budget, investment, and commissioning cycles.

Good practice is to base value assessments on credible evidence of outcomes and impact, using methods that best fit the evaluation question rather than a single prescribed approach, this includes:

- being transparent about assumptions, costs and benefits
- involving partners and stakeholders to identify outcomes that matter
- ensuring outcomes are measured or have a credible plan for measurement.

It also involves:

- considering how costs and benefits are distributed across different groups
- being explicit about timeframes and assumptions (including discounting)
- using recognised costing and valuation approaches where appropriate
- avoiding overclaiming by only attributing value that is reasonably supported by the available evidence.