

## About Superu

superu

Superu's purpose is to increase the use of evidence by people across the social sector so that they can make better decisions – about funding, policies or services – to improve the lives of New Zealanders, New Zealand communities, families and whānau.

The evidence rating scale is one stream of work that sits under Superu's *Using Evidence for Impact* project. This project drives:

- greater accessibility to evidence
- greater transparency of evidence
- capability development and good practice in using evidence.

Social Policy Evaluation and Research Unit

superu

## An evidence rating scale for New Zealand

### Understanding the effectiveness of interventions in the social sector

*Using Evidence for Impact*

JUNE 2017



The Families Commission operates under the name  
Social Policy Evaluation and Research Unit (Superu)

ISBN: 978-0-947489-82-3 (print) ISBN: 978-0-947489-83-0 (online)

Learn more at: [superu.govt.nz](http://superu.govt.nz)

*Quick reference guide*



## Why do we need a rating scale?

---

Government agencies and the service providers they invest in undertake a substantial amount of research and evaluation. Such evidence, when of the right quality, is a crucial part of decision making around whether or not the programmes or services they develop, invest in or deliver make a positive difference to vulnerable New Zealanders.

**This rating scale** provides a standard against which the evidence for effectiveness of social sector interventions can be assessed. It enables:

- a consistent and transparent approach to assessing whether a programme or service has sufficient evidence about outcomes, or good evidence for positive, nil, or negative outcomes
- an explicit mechanism for evidence to influence decisions around investment (including scalability and implementation in new locations)
- increased visibility of evaluation as a decision making tool
- better quality evaluation, more use of evidence in decision making, and more sharing of findings about what programmes or services work.



## What is the rating scale?

---

The scale provides a pathway to excellence, and a set of criteria against which evidence can be assessed. **Ultimately it will lead to the delivery of more effective and efficient programmes and services.**

The two dimensional scale looks at both the strength of evidence for an intervention (taking into account the maturity or otherwise of a programme or service), as well as its effectiveness (if it exists) – be that beneficial or harmful – to both the target population and to others. The scale has been designed to be inclusive of different evaluation approaches (western or Māori).

The full rating scale differentiates between New Zealand and overseas evidence, demanding more from overseas so that transferability (i.e. is the intervention a good fit for New Zealand) and issues of scale are properly assessed.

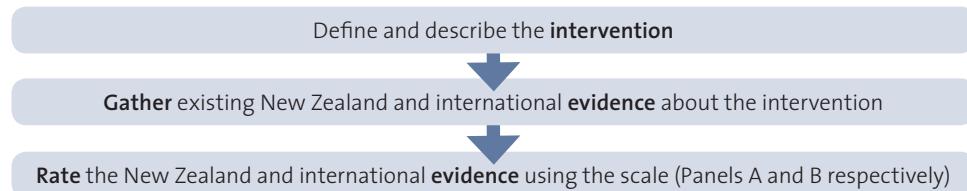
The scale does not judge the quality of individual pieces of research (although this will need to be done as part of the process), nor does it set out to judge the value of research on a specific sector.

### More information

- The full document – *An evidence rating scale for New Zealand: Understanding the effectiveness of interventions in the social sector* can be found at [superu.govt.nz/resources/evidence-rating-scale](https://superu.govt.nz/resources/evidence-rating-scale)
- For help with the language, see [superu.govt.nz/resources/glossary](https://superu.govt.nz/resources/glossary)

## How do you use the rating scale?

The scale can be used by a range of people or organisations, from small community or non-government organisations wanting to provide a rationale for funding or to build their evidence effectiveness, through to central government agencies designing and funding large-scale investments in the social sector. It involves the following steps:



For more detail on using criteria see the full document at: [superu.govt.nz/resources/evidence-rating-scale](https://superu.govt.nz/resources/evidence-rating-scale)

## The effectiveness scale

The **vertical scale** assesses the effectiveness of an intervention as it is important to understand the effectiveness – be that harmful or beneficial – to both the target population and to others who might be affected.

Criteria		At levels
<b>Beneficial</b>	Strength of evidence demonstrates positive effects on desired outcomes and shows no substantive harmful effects of the intervention.	2 to 4
<b>Mixed effects</b>	Some evaluation/s demonstrate positive effects while others demonstrate no effects on any desired outcomes, and there has been no demonstration of substantive harmful effects of the intervention.	2 to 4
<b>No effect</b>	Strength of evidence demonstrates no positive effects on any desired outcomes and shows no substantive harmful effects of the intervention.	2 to 4
<b>Harmful</b>	Strength of evidence shows substantive harmful effects of the intervention on the target population, or has harmful effects on others that outweigh the benefits to the target population.	2 to 4
<b>Not applicable</b>	Strength of evidence is insufficient to assess effectiveness.	0 to 1

## Panel A: Should we fund or continue this New Zealand intervention?

	0 Pilot initiative	1 Early stage, good in theory	2 Progressing, some evidence	3 Good evidence, sufficient for most interventions	4 Extra evidence for large or high risk interventions
<b>Beneficial</b>	✓	✓	✓	✓	✓
<b>Mixed</b>	Strong theory of change with evidence-based logic	Too soon for effectiveness data, but processes and data suggest it is on track	✓	Consider weight of evidence, risk, alternatives	Consider weight of evidence, risk, alternatives
<b>No effect</b>			Consider stopping	✗	✗
<b>Harmful</b>			✗	✗	✗

## Panel B: Should we consider implementing this intervention in New Zealand?

	0 Pilot initiative	1 Early stage, good in theory	2 Progressing, some evidence	3 Good evidence	4 Well supported, large scale
<b>Beneficial</b>	✗	✗	Maybe depending on alternatives	Maybe depending on alternatives	✓
<b>Mixed</b>	No effectiveness data yet	No effectiveness data yet	✗	Consider weight of evidence, risk, alternatives	Consider weight of evidence, risk, alternatives
<b>No effect</b>			✗	✗	✗
<b>Harmful</b>			✗	✗	✗

## The strength of evidence scale

The **horizontal scale** relates to the overall strength of a body of evidence. This scale consists of five levels (0 to 4), taking into account both the strength of evidence, and expectations about the type of evidence that can and should be generated about an intervention as it matures and grows.

This scale should be able to judge the strength of evidence from any evaluation and also includes intervention consistency and documentation. As shown in the table immediately below, efficiency, effectiveness and impact are among the key things that need to be addressed by an evaluation.

Criteria	Level				
	0	1	2	3	4
Evidence-based logic model or theory of change	✓	✓	✓	✓	✓
Evaluation plan	✓	✓	✓	✓	✓
Minimum number of evaluations of the characteristics shown in the corresponding level of the table to the right			1	1	2
Minimum number of implementations/replications					2
<b>Intervention consistency and documentation criteria</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
Comprehensive documentation			✓	✓	✓
Procedures are in place to ensure consistent implementation			✓	✓	✓
Information is available on resources required to deliver			✓	✓	✓
Regular review/quality assurance procedures				✓	✓
Technical support is available to help implement in new settings					✓

The available evaluation(s) should...	For level		
	2	3	4
– report on <b>efficiency</b>	✓	✓	✓
– assess <b>effectiveness</b>	✓	✓	Strongly/reaching target groups
– assess <b>cost-effectiveness</b>		✓	Cost-benefit analysis
– indicate <b>impact</b>	Some	Attribution	Strongly/on sub-groups
– identify <b>mechanisms</b>		Causal	Modifiable
– provide guidance for implementation in <b>new settings</b>		Some	Supports
– used appropriate/ <b>robust methods</b> to measure change	✓	Pre- & post-analysis	Pre- & post-analysis
– used appropriate <b>analysis</b> /has sound conclusions	✓	✓	✓