Evaluation Report Assessment Tool
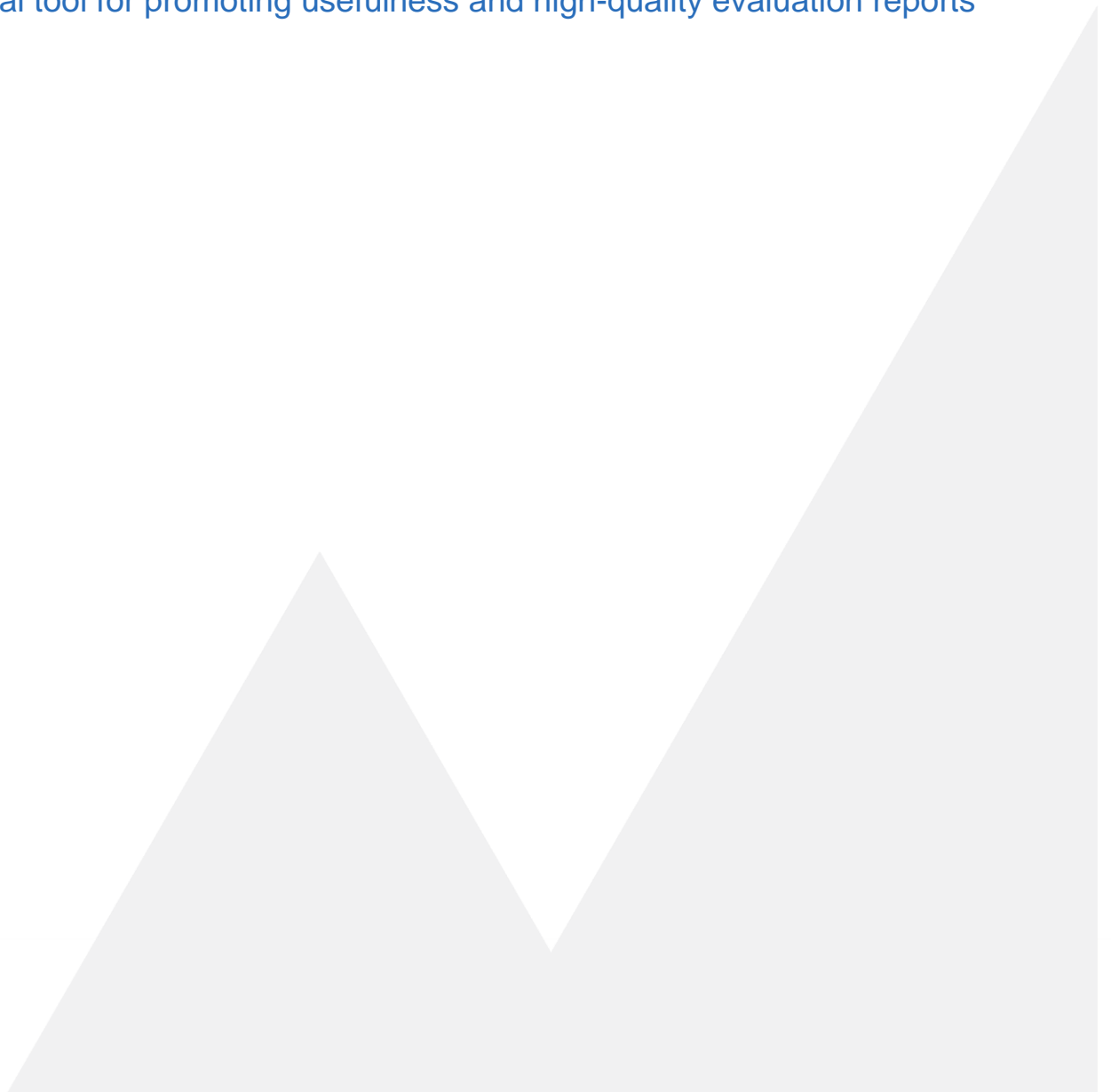
A practical tool for promoting usefulness and high-quality evaluation reports
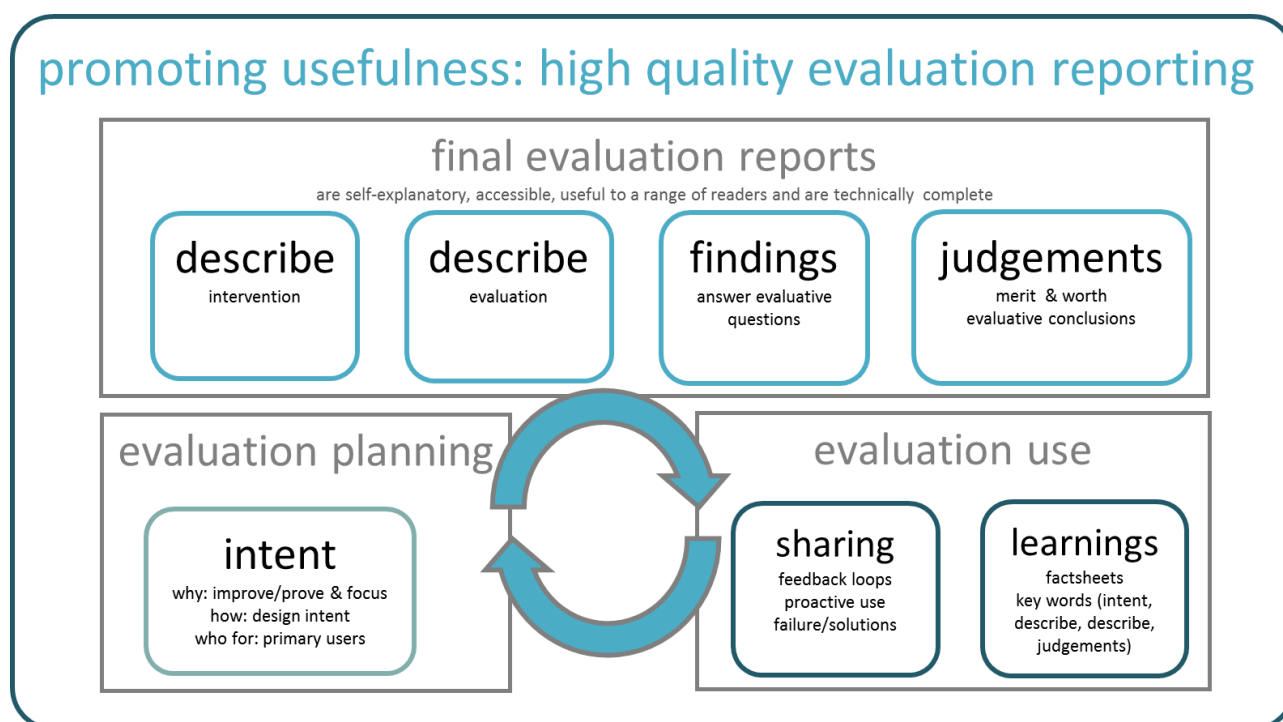
Acknowledgments

Contents

Evaluation Reports

The Ministry of Education commissions a number of evaluations each year. Evaluations have an important role to influence the development of the system, improve decision-making and inform the Ministry how well it is delivering on strategic priorities.

It is essential to the Ministry that evaluation reports are **high-quality** and **useful**[1]. The constructs of 'high-quality' and 'usefulness', and promoting these, have been conceptualised in the logic diagram and text description below (see Figure 1).

Figure 1 Evaluation reports require planning, the right content and proactive use to be high quality and useful

High quality evaluation reports are **accessible**, have the right **content** (i.e. **describe** the intervention and evaluation, present **findings** and **judgements**) spread across **summaries** and **chapters** with details in rich **appendices.** They provide **coverage** across the breadth and levels of Ministry activities.

promoting usefulness: high quality evaluation reporting

final evaluation reports
are self-explanatory, accessible, useful to a range of readers and are technically complete

describe
intervention

describe
evaluation

findings
answer evaluative questions

judgements
merit & worth
evaluative conclusions

evaluation planning

intent
why: improve/prove & focus
how: design intent
who for: primary users

evaluation use

sharing
feedback loops
proactive use
failure/solutions

learnings
factsheets
key words (intent, describe, describe, judgements)

Enhanced usefulness of evaluation reporting will be achieved when high quality evaluation reports are produced from careful **evaluation planning** and proactive **evaluation use.**

---

[1] See full report for more details: Dowden, A. (2018) *Meta-Evaluation report. Promoting usefulness: Evaluation Reports. Using an assessment framework to assess final Evaluation Reports at Ministry of Education.* Anne Dowden REWA. Wellington.

Evaluation planning

Intent of the evaluation (why, how, who for).

- **why** the Ministry is doing an evaluation: whether it aims to prove, improve, or monitor etc;
- **how** to do the evaluation: design options;
- **who** the evaluation is for: audiences, users and uses of the evaluation.

Evaluation report content

Note: Key details should be included in report summaries while full details should be in **appendices**.

Describe

Describe the intervention – what programme/policy/strategy the evaluation will focus on, its strategic context; what its intention is (problem definition and intervention aim/s), and what its core actions are.

Describe

Describe the evaluation – scope, evaluation questions, methods (research type methods – data collection/collation/analyses approaches; evaluative methods - evaluation framework, logic diagrams, success indicators/rubrics, processes to make judgements).

Findings

The **findings** should **systematically** present **synthesised data** to answer **all evaluation questions succinctly**, with **adequate justification**, with use of relevant **graphs and tables** to illustrate arguments, with inclusion of some commentary on **limitations.**

Reports should provide **summative statements** about the findings that linked back to the evidence (these can be included with the findings or with the conclusions).

- **Synthesising data** is a sensemaking process of combining and presenting evidence (data/information) from all relevant sources (e.g. collated administrative data, collected primary data, literature, previous evaluations and research). Synthesis is achieved by the process of weighing up which evidence is more or less important and considering, on balance, how well the evidence informs the area of investigation posed by each evaluation question. Several pieces of evidence would usually be used to provide adequate justification for an argument. The evidence presented should be specific and relevant to the area of investigation, rather than a 'dump' of all available evidence. Appendices can be used to present useful, but less relevant, evidence. The synthesised evidence that is presented should be carefully selected, and the reasons for selection need to be defensible (peer review or stakeholder workshopping are useful to assess defensibility of your process).
- **Summative statements** about the findings should be clear and specific. These statements should summarise the findings and one or more statements should focus on the area of investigation posed by each evaluation question. The statements should overtly draw on supporting key evidence (data/information) to **justify** the point made. Use of cross-references are an effective and user-friendly technique to use.
- All questions should be answered. If no evidence is available for one or more evaluation questions, report this and explain why.

Conclusions

Conclusions are made up of:

- Judgments - making a judgement is the process of using data and information to make judgements about the success (or otherwise) of an intervention.
- Implications – commentary on the implications of the judgements made about the intervention, i.e. what the judgements mean for the intervention within its context.
- Considerations or Recommendations – what next - what to do now that we have the judgements and understand the implications of these.

Judgements should be **clear**, **unambiguous statements** on the merit or worth of the area of investigation posed by each evaluation question, or part of the intervention, or the entire intervention.

- A judgment should state **how well/how successful[2]**, the achievements/the change/the quality of implementation (and so forth) has been met, and should be related to **agreed definitions of 'success'** (these may be written and formalised, and be in place prior to evaluation; or they may be less formal consensus-based definitions developed during the evaluation).
    - A judgement might state that achievement was 'at' or 'above' the expectation, the target level, a minimum safety standard, a quality standard, or is higher than the counterfactual, or meets or succeeds the success criteria or definition in a rubric.
    - The judgement process might be achieved during a findings workshop of key stakeholders who bring their expert views to assess the findings collectively to confirm how well a constellation of evidence represents being 'at' 'above' or 'below' a commonly held views of what 'success' looks like.
    - The judgment process might be based on a combination of these approaches.
- The judgement should be **traceable back to the evidence** presented in the findings. A user-friendly technique to provide traceability is briefly summarise and cross-reference key evidence. **Frameworks that map** evidence to each evaluation question and to rubrics or success indicators are useful for this.

Considerations or **Recommendations.** Deciding whether or not an evaluation ought to overtly and formally 'recommend' actions is a complex decision.

- Formal **recommendations** should either be traceable to evidence in the report findings or overtly be an 'expert' recommendation. If expert opinion is offered, then the expertise should be clearly demonstrated in the report. Expertise is a somewhat subjective and intangible construct. Recommendations based on a combination of traceability and expertise are eminently more acceptable to a wider range of audiences.
- **Considerations** or 'Next steps' offer a vehicle to present a set of actions, key messages, or investigations to consider that are based on analysis of the evidence in its context. These are useful but not so overtly instructive as formal recommendations. Considerations or Next steps serve a useful role of offering advice that can be considered alongside a range of other information (such as political acceptability, budget constraints and so forth which may also impact on decisions).

---

[2] They should <u>not</u> simple repeat the findings or the summary of findings "50% of people said xyz", without providing a judgement about how 'good' the finding is. While this statement appears to show the intervention is a 'success', if the progress was below expectation then such an interpretation would be wrong.

Evaluation use

A key step to promoting **evaluation use** is for evaluation reports to be user friendly: they should be **actionable, navigable, accessible,** and **complete.**

Actionable reporting[3] has a focus on making the front end (the executive summary) of a report as easy to digest and as complete and standalone as possible – so that the reader has all the key information to immediately 'take action'. This includes content and format as follows:

- Re-ordering the traditional scientific reporting model.
- Uses 25-page summary report format (or a 1, 3, 25-page format).
- Presenting finding summaries and conclusions together – which enhances traceability.

Table 1. Actionable reporting

Action-orientated report order
1. Cover
2. Executive Summary
3. Condensed Front Matter
4. Findings & Interpretation
5. Appendices: Methods & Analysis, Design, Background

---

[3] Source: Ewing Marion Kauffman Foundation. *Evaluation Report Guidance*, November 2017. https://www.kauffman.org/-/media/kauffman_org/evaluation/emkf-evaluation-report-guide-2017.pdf?la=en

Development of Assessment Tool

Tool development

The **Evaluation Report Assessment Tool** was developed using an evidence base gathered from international literature[4]. This evidence was augmented by Ministry contextual information (including the Ministry's strategic direction and ways of working). The tool links to the conceptual logic that shows how the Ministry wants to promote usefulness and high-quality evaluation reports (see Figure 1). The tool covers six areas, see Table 2, below.

Table 2. Evidence base for Assessment Tool

Evaluation Report Assessment Tool	Links to report logic 'promoting usefulness' (Figure 1)
1. User focused actionable, navigable, accessible, complete	evaluation use
2. Front title page, content, acknowledgements etc.	evaluation use describe
3. Summary intervention, context, evaluation questions, methods, findings, judgements	describe findings judgements
4. Findings chapters succinct presentation of findings for all evaluation questions	findings
5. Conclusions judgements, their implications, considerations/recommendations	judgements
6. Appendices cross referenced detailed findings, details of intervention/evaluation/methodologies	describe findings

A finer look at the appendices

The evaluation report appendices can be very large and contain a great deal of information (especially when the majority of the methodology and details of the analysis are included in the appendices). For a more straightforward assessment, first **consider rating presence of *Essential* information**. This may be an especially useful approach if the appendices appear to be well below standard. Essential details include:

- Self-explanatory description of the intervention
- Evaluation intent (what's its purpose)
- Some evaluation questions are evaluative (not only descriptive)
- Method mapping: sources/data mapped to all Evaluation questions/sub questions
- Judgement tool(s) / procedure(s) used to assess evidence to determine worth (to judge)
- Information needs (i.e. to answer the evaluation questions)
- Sources of data/information
- Limitations (issues, reliability, validity, use of proxy)
- No glaring *evaluation methodology* quality issues.

Notwithstanding the above suggestion, it is important to be aware that high-quality and useful reports should include *Important* and *Useful* information. See **over page** for details identified as *Essential, Important* and so forth.

---

[4] See full report (Dowden, 2018) for more details.

Figure 2 Key for detailed assessment of appendices

Essential	
Important	
Should be included or available (e.g. in a separate Evaluation Plan)	
Useful extra detail (consider including)	

Report Appendices

Cross referencing	☐ All appendices referred to report are included
Bibliography /references	☐ Included with report
	☐ Complete, standardised/acceptable format
Appendices	☐ Title
	☐ Source / referenced
	☐ Clearly presented, complete, self-explanatory
Methodology	
Intervention logic high level overview	☐ Context: within education system; MoE Strategy and Operating Model
	☐ Self-explanatory description
	☐ Provides overview without excessive in detail
Evaluation reason and scope	☐ Evaluation intent (what's its purpose)
	☐ Evaluation uses, & users (audiences)
	☐ Evaluation level (intervention/ system/ strategic)
	☐ OPTIONAL evaluation approach/philosophy
	☐ OPTIONAL focused logic showing evaluation scope, focus, and relevant intervention outcomes
	☐ What is in (and out of) scope
	☐ In scope/relevant intervention outcomes listed
	☐ OPTIONAL rationale for scope
Evaluation framework (and evaluation method)	☐ Evaluation questions linked to intervention outcomes
	☐ Some questions are evaluative (not only descriptive)
	☐ Sub questions 'for whom', 'in what circumstances'
	☐ Method framework mapping: sources/data mapped to all Evaluation questions/sub questions
	☐ Judgement tool(s)/procedure(s) used to assess evidence to determine worth (to judge), criteria, indicators, rubrics (& is linked to evidence base)
	☐ Evaluation team expertise if using 'expert judgement'
Summary list methods	☐ All methods included
Methodology (technical procedure for each method to gather evidence)	☐ Information needs
	☐ Sources
	☐ Ethics: consent, confidentiality
	☐ Cultural considerations
	☐ Describe/attach collection tools & processes
	☐ Limitations (issues, reliability, validity, use of proxy)
Method quality (esp. considering evaluation intent & resources)	☐ No glaring *evaluation methodology* quality issues
	☐ No glaring quality issues in *'research'* techniques.
	☐ Any quality issues are fully disclosed and managed

Evaluation Report Assessment Tool

Grading evaluation reports using the Tool

Each item on the assessment tool can be given a grade, using the following symbols.

- ☑ Present / high quality, meets the standard, or meets needs
- ⊟ Present / low quality, inconsistent, or partially meets needs
- ☒ Not found, no evidence of presence, below standard, or does not meet needs
- ⊡ NA = not applicable

Note: Where content is not in an appendix, but is in other sections of the report it should still be rated e.g. background or method details may be in an introduction or a methodology section. In this instance, feedback about the report might include suggestions to move extensive details to appendices so that the report is more **actionable** in format.

Four-page Evaluation Report Assessment Tool

See over page, copies can be download by clicking in this link: Evaluation Report Assessment Tool October 2018 FINAL.pdf

Support for using the Tool and ensuring its usefulness

Evaluation Capability work of the Ministry will include training on doing report assessments and feeding back to report authors, as well as setting up mechanisms for peer review of completed report assessments.

The Tool will be a living document that evolves based on feedback from users. EDK will welcome user-feedback on how the Tool could be improved to better meet users' needs.

Evaluation Report Assessment Tool
developed by **Anne Dowden REWA**

MINISTRY OF EDUCATION
TE TĀHUHU O TE MĀTAURANGA

Report Title: _____

Review Date: _____   Reviewer: _____

Peer Reviewer: _____

Key to symbols for grading reports.

☑ Present / high quality, meets the standard, or meets needs
⊟ Present /low quality, inconsistent, or partially meets needs
☒ Not found, no evidence, or not meet needs
⊡ NA = not applicable

Tool Table A: Assessment tool user focused: actionable, navigable and accessible

User focused – actionable, navigable, accessible, complete		
Actionable report format	☐	25pg: Title, Contents, Summary, Findings, Conclusions. Appendices (background, design & methodology, tabulations, detailed analyses etc.)
Navigable	☐	Contents pages/cross refs have active hyperlinks
	☐	Main headings sequentially marked (numbered)
	☐	Appendices use own page numbers (e.g. A-1, A22)
	☐	PLUS Navigation aids (mini-logics, colours)
Accessible		
Accuracy, clear	☐	Accurate (error free)
	☐	Clear meaning (not confusing or repetitive)
Accessible to non-specialist educated reader	☐	Accurate non-technical summaries
	☐	Meaningful skim reading – due to subtitles/header paragraphs etc.
	☐	Detailed analyses are succinct with limited jargon
Complete	☐	Does not assume prior knowledge (e.g. of intervention or previous evaluations)
Chapters accessibility		
Chapter contents	☐	Note about chapter - what it's presenting
Chapter summaries	☐	Summary of chapter findings (I.E. ARE PRESENT)
Headings - chapter, subsections & charts	☐	Informative headings used throughout "Students' experience successes" not "Students' stories"
Reference to sources/explanations	☐	(PLUS for access) Cross refs, summary of literature Cross refs, appendices of tabulations Cross refs, detailed explanations in other chapters
Format of all features (headings, text, charts, tabulations, diagrams)	☐	Font type, ideally Calibri 11pt – or equivalents e.g. Arial 10pt (or variants narrow/**bold**)
	☐	Size consistent, minimum 10pt narrow in any item
	☐	Formatting is consistent
	☐	Sub headings styles consistent, stand out
Graphs and data tables consistently show	☐	Title
	☐	Data/axis/columns labelled
	☐	Sample size
	☐	Source / data collection tool (e.g. student survey)
	☐	PLUS May footnote data limitations

Notes about any concerns/failures/feedback

Tool Table B: Assessment tool: front and summary

Report Content - Front		
Title page	☐	Clear title (project, evaluation type)
	☐	Client (name and organisation)
	☐	Date
	☐	Correct use of the term evaluation
Evaluation team	☐	Contact details (at least organisation name)
	☐	Team affiliations, expertise, roles
	☐	Acknowledgments: data collectors, community
Contents page	☐	Titles, subtitles, appendices, page numbers
	☐	OPTIONAL List of charts and tables
Glossary	☐	All technical terms & acronyms
Report Content - Summary (e.g. report/executive summary)		
Intervention	☐	Problem definition
	☐	Intent/aim
	☐	Intervention description - core activities
Intervention in context	☐	Intervention within system (ECE, schooling, tertiary) *
	☐	Location within Ministry strategic priorities
	☐	Intervention type (e.g. behaviour management)
Evaluation intent & questions	☐	Evaluation purpose (monitor, prove, improve, review)
	☐	List of evaluation questions
	☐	Cross reference to full evaluation framework
Evaluation methodology (short format)	☐	List of data collection method activities
	☐	Short summary of how evaluative judgements made
	☐	Cross reference to evaluation focus and scope/ evaluation framework/methods
Findings	☐	Findings for all evaluation questions
	☐	Presents key evidence for each finding
	☐	Cross reference to detailed findings
Summative judgement/s	☐	Judgement on <u>all</u> evaluation question about success of Intervention
	☐	Clear unambiguous statements (YES/NOT ONLY)

Notes about any concerns/failures

Tool Table C: Assessment tool: findings and conclusions

Report content – Findings		
Method & design chapter		*(see appendix for assessment)*
Answer evaluation questions	☐	Reports by question; not by data tool or when/how collected
Present all the central evidence to answer the question	☐	Succinctly presents enough evidence to answer evaluation question and to make a credible finding.
	☐	Findings in chapter are adequately justified
	☐	Cross references to contributing evidence (e.g. previous analysis, tabulations, literature)
Graphs and data tables	☐	Clear which data is being shown
	☐	Relevant to argument
(PLUS adds strength): Alternative explanations	☐	Evidence of causality, pre-post change or control groups, describes the counterfactual (what would happen without the intervention)
(PLUS adds strength): Puzzles and unexpected outcomes	☐	Describes outcomes that are unintended, unexplained, emergent, or negative
	☐	Comments on their cause, and whether or not they are harmful or advantageous (and to whom)
Commentary on strength/limitations of the data	☐	Data quality and validity; whether it can be used (is it trustworthy enough in these circumstances)
All questions	☐	All evaluation questions answered in report (or reasons why not are given)
Summative statements	☐	Clearly states findings for all evaluation questions
	☐	Findings are linked to evidence
	☐	Findings in summary are adequately justified
	☐	Commentary on how strong the findings are

Report content - Conclusions		
Judgements *"So what?"*	☐	Judgements of merit/worth are included
	☐	Judgements have clear, traceable links to findings
Implications *"Now what?"*	☐	Commentary on implications of judgements
Considerations *"What next?"*	☐	Next steps commentary (e.g. intervention challenges /development, new evaluation foci)
Recommendations *"What next?"*	☐	Recommendations are based on report findings
	☐	Recommendations based on overtly referenced expertise

Notes about any concerns/failures

Evaluation Report Assessment Tool
developed by **Anne Dowden REWA**

Tool Table D: Assessment tool: appendices methodologies and data

Appendices		
Cross referencing	☐	All appendices referred to report are included
Bibliography/reference list	☐ ☐	Included with report Complete, standardised/acceptable format
Appendices (see Table 1 to assess content of graphs / data tables)	☐ ☐ ☐	Title Source / referenced Clearly presented, complete, self-explanatory
Methodology		
Intervention logic high level overview presenting intervention in context	☐ ☐ ☐	Context: within education system; MoE Strategy and Operating Model Self-explanatory Provides overview without excessive in detail
Evaluation reason and scope	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	Evaluation intent (what's its purpose) Evaluation uses, & users (audiences) Evaluation level (intervention/ system/ strategic) OPTIONAL evaluation approach/philosophy What is in (and out of) scope OPTIONAL focused logic showing evaluation scope, focus, and relevant intervention outcomes In scope/relevant intervention outcomes listed OPTIONAL rationale for scope
Evaluation framework (and evaluation method)	☐ ☐ ☐ ☐ ☐ ☐	Evaluation questions linked to intervention outcomes Some questions are evaluative (not only descriptive) Sub questions 'for whom', 'in what circumstances' Method framework mapping: sources/data mapped to all Evaluation questions/sub questions Judgement tool(s) / procedure(s) used to assess evidence to determine worth (to judge), criteria, indicators, rubrics (& is linked to evidence base) Evaluation team expertise if using 'expert judgement'
Summary list methods to gather evidence (mid length format and non-technical)	☐	All methods included
Methodology (technical procedure for each method to gather evidence)	☐ ☐ ☐ ☐ ☐ ☐	Information needs Sources Ethics: consent, confidentiality Cultural considerations Describe/attach collection tools & processes Limitations (issues, reliability, validity, use of proxy)
Method quality (consider evaluation intent and resources - such as time and budget; consider nature/size of decisions to be made) ☑ **IF AGREE NO GLARING ISSUES** ☒ **IF DISAGREE AND SEE GLARING ISSUES**	☐ ☐ ☐	No glaring *evaluation methodology* quality issues (i.e. evaluation scope, evaluation questions, evaluative framework, findings synthesis, evaluative judgements, implications/recommendations) No glaring quality issues in *'research' techniques.* Any quality issues are fully disclosed and managed so that inaccurate conclusions or judgements are not made solely using poor quality evidence / processes

We **shape** an **education** system that delivers **equitable** and **excellent outcomes**

He mea **tārai** e mātou te **mātauranga** kia **rangatira** ai, kia **mana taurite** ai ōna **huanga**