

Standards of evidence



The *Standards of evidence* help AERO and the education community make consistent and transparent judgements when assessing education evidence.

Why do we need standards of evidence?

Education practitioners and policymakers make dozens of decisions every day aimed at improving outcomes for their children and students.

How can they be confident that what they choose to do will work? How do they decide between one approach versus another?

To answer these questions in a consistent and transparent way, education practitioners and policymakers need a way to evaluate the strength of research evidence on the effectiveness of a particular approach. They need standards of evidence.

What are AERO's Standards of evidence?

AERO's *Standards of evidence* establish AERO's view on what constitutes *rigorous* and *relevant* evidence. When evidence is rigorous and relevant, it provides confidence that a particular approach is effective in a particular context.

AERO's *Standards of evidence* can apply to all forms of education evidence – whether generated through academic research or by education practitioners through their daily practice.

AERO will primarily use the *Standards of evidence* when undertaking those projects which are based on syntheses and causal (or evaluative) research.

In developing the *Standards of evidence*, AERO has sought to build upon existing policies and research on evidence use across Australia and the world. AERO hopes that the *Standards of evidence* and associated evidence tools can help enhance quality conversations about generating and using evidence across the Australian education community.

What makes evidence rigorous and relevant?

There are many criteria that can be used to evaluate education evidence. AERO's *Standards of evidence* prioritise two criteria: *rigour* and *relevance*. These criteria have been prioritised because they are the most important considerations when deciding whether a piece of evidence can give someone confidence that a particular educational approach will be effective in their context.

Rigorous evidence is defined as evidence produced using research methods (whether qualitative, quantitative or mixed methods) that isolate the specific impact of a particular educational approach.

Relevant evidence is defined as evidence produced in contexts that are similar to one's own. Evidence is also relevant when it is derived from a large number of studies conducted over a wide range of contexts, as this suggests that the educational approach is not dependent on any particular contextual factor.

Although the *Standards of evidence* clearly differentiate between four levels of confidence that evidence can provide, the standards should be viewed as a continuum along which rigour and relevance gradually increase. Evidence at each level builds on the evidence from preceding levels.

How does AERO use the *Standards of evidence*?

AERO uses the *Standards of evidence* in two ways. When conducting research syntheses, AERO uses the *Standards of evidence* to make consistent and transparent judgements in selecting the best available research evidence. When conducting causal (or evaluative) research, AERO uses the *Standards of evidence* to guide the design and implementation of the research, so that it generates high-quality evidence for the Australian education community.

There may be occasions when research evidence does not clearly fit within a particular level of confidence. When this is the case, AERO draws on expert research guidance to make an assessment about how confident we are in the effectiveness of an approach.

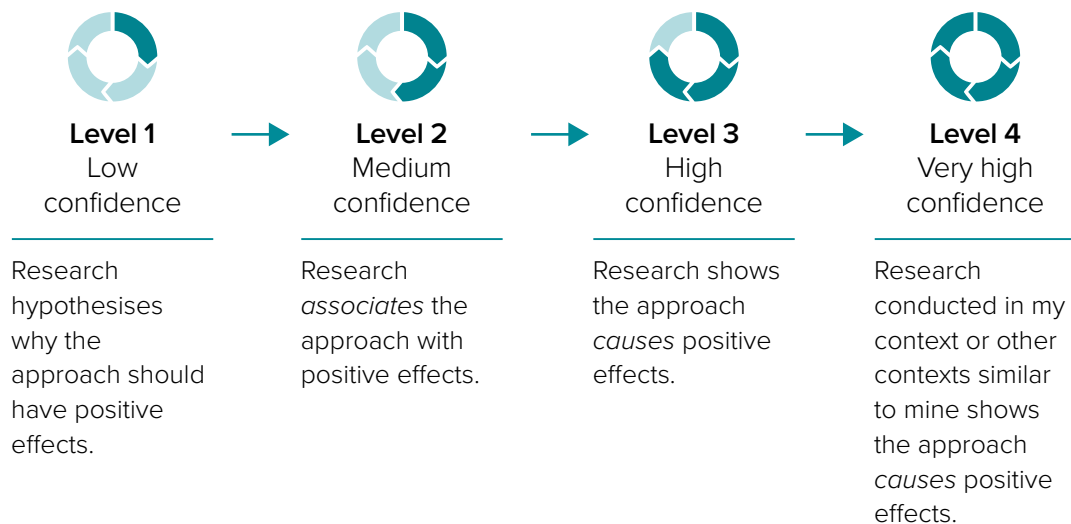
How can education practitioners, policymakers and researchers use the *Standards of evidence*?

The *Standards of evidence* can be used by education practitioners, policymakers and researchers interested in determining the strength of existing evidence for a particular approach in their particular context.

To help practitioners and policymakers use the *Standards of evidence* in their context, AERO has developed [Evidence rubrics](#). The rubrics help practitioners and policymakers evaluate their confidence in the effectiveness of a new or existing approach and provide implementation guidance appropriate to one's level of confidence.

The *Standards of evidence* can also be used by policymakers and researchers when designing evaluations. By designing evaluations aligned to the *Standards of evidence*, policymakers and researchers can try to generate evidence that meets a desired level of confidence.

Standards of evidence





Level 1 Low confidence

Research hypothesises why the approach should have positive effects.

<p>What types of research fit within this level?</p>	<p>Research that presents a hypothesis for why the approach should have positive effects on outcomes.</p> <p>This research does not provide data (whether qualitative or quantitative) to substantiate its claims that the approach is effective.</p>
<p>What features of research studies increase my confidence within this level?</p>	<p>The study provides an explanation that is based on well-established theories of learning and development.</p> <p>The study clearly explains step-by-step how the approach is hypothesised to have positive effects.</p>



Level 2 Medium confidence

Research *associates* the approach with positive effects.

<p>What types of research fit within this level?</p>	<p>Research that demonstrates a correlation between the approach and positive effects on outcomes; for example:</p> <ul style="list-style-type: none"> • small-scale studies, such as case studies, and/or large-scale studies, such as cross-national surveys • studies using qualitative (e.g. observations and/or interviews), quantitative (e.g. statistical techniques) or mixed methods. <p>This research does not necessarily show that the approach causes positive effects as there could be other potential explanations.</p>
<p>What features of research studies increase my confidence within this level?</p>	<p>The study has been conducted in my own context or in contexts similar to my own.</p> <p>The study corroborates findings from other studies conducted in many different contexts.</p> <p>The study measures change in outcomes over time.</p> <p>The study has a large sample size that is spread across more than one site.</p> <p>The study uses strategies that discount the possibility that effects are due to chance.</p> <p>The study compares one group that has been subject to the approach to another group that has not been subject to the approach.</p> <p>The study is conducted by people or organisations independent of the developer of the approach.</p> <p>The study has been conducted recently.</p>



Level 3 High confidence

Research shows the approach *causes* positive effects.

<p>What types of research fit within this level?</p>	<p>Research that meets the following criteria:</p> <ul style="list-style-type: none"> • uses rigorous qualitative, quantitative or mixed methods that address issues like selection bias, history effects and maturation effects • uses outcome measures validated for the purposes of the study. <p>This research does not necessarily prove the approach causes positive effects <i>in my context</i>. This is because there may be other factors in my context that mean the approach will not work as intended.</p>
<p>What features of research studies increase my confidence within this level?</p>	<p>The study corroborates findings from other studies conducted in many different contexts.</p> <p>The study measures change in outcomes over time.</p> <p>The study has a large sample size that is spread across more than one site.</p> <p>The study uses strategies that discount the possibility that effects are due to chance.</p> <p>The study compares one group that has been subject to the approach to another group that has not been subject to the approach.</p> <p>The study has been conducted by people or organisations independent of the developer of the approach.</p> <p>The study has been conducted recently.</p> <p>The study mitigates the likelihood that effects are simply due to the particular characteristics of those that participate in the study.</p> <p>The study discusses and/or tests the key contextual factors that may influence the effectiveness of the approach.</p>



Level 4 Very high confidence

Research conducted in my context or contexts similar to mine shows the approach *causes* positive effects.

<p>What types of research fit within this level?</p>	<p>Research that meets the following criteria:</p> <ul style="list-style-type: none"> • uses rigorous qualitative, quantitative or mixed methods that address concerns like selection bias, history effects and maturation effects • uses outcome measures validated for the purposes of the study • is conducted in my context or in contexts similar to mine.
<p>What features of research studies increase my confidence within this level?</p>	<p>The study corroborates findings from other studies conducted in many different contexts.</p> <p>The study identifies the factors that lead to the approach working, and the conditions that are necessary for the approach to be implemented on a larger scale.</p> <p>The study assesses the effectiveness of the approach on different subgroups and explains reasons for any differences in effectiveness between subgroups.</p> <p>The study monitors outcomes for different groups over time to ensure continued effectiveness.</p>