Practice guide for ECEC and schools

Build professional knowledge

Australian Education Research Organisation

Assessing research evidence

Using research to strengthen your practice

March 2024

Knowing how to assess the reliability and relevance of research evidence can help you make better decisions for adopting and implementing new practices.

As an education professional, drawing on research evidence can strengthen your confidence that you're using the most effective practices – practices that will maximise children and young people's learning and wellbeing. Considering new approaches doesn't mean your existing approaches 'don't work'. It's simply recognising that knowledge evolves over time. Even when children and young people are doing well, if we ignore the research, they could miss out on opportunities to achieve more. Research evidence shouldn't replace the professional wisdom and understanding that comes with experience. Rather, it should complement your experience, enrich your existing knowledge and give you confidence you're using the practices most likely to maximise learning in your context.

The Australian Education Research Organisation (AERO) has produced 4 practice guides on using research to strengthen your practice:

- The Value of Research Evidence
- Assessing Research Evidence
- Looking for Research Evidence
- Applying Research Evidence.

If you're a teacher or educator, this practice guide and AERO's other evidence use resources can help you draw effectively on research evidence to strengthen decisions about your practice. If you're a school or service leader, you can use this guide and these resources to support your team in engaging with research evidence as part of their ongoing professional development.

Related frameworks

Early Years Learning Framework V2.0

Principles: Critical reflection and ongoing professional learning.

National Quality Standards

Standard 7.2 Leadership: Effective leadership builds and promotes a positive organisational culture and professional learning community.

Australian Professional Standards for Teachers

Focus Area 6.2: Engage in professional learning and improve practice, which includes 'Plan for professional learning by accessing and critiquing relevant research' at the Highly Accomplished level.

Australian Professional Standards for Principals

Professional Practice 2: Developing self and others.

Ways to use this practice guide

- You can use this practice guide for professional learning to become more familiar with research and to check your knowledge.
- You can use this practice guide for professional learning to discuss research evidence as a team, such as in a community of practice.
- Leaders can use this practice guide to structure dialogue and reflection about using research evidence in a school or service. These concepts can serve as a point of conversation to build shared understandings of how to engage with research evidence.

Assess research before you rely on it to make a decision

Regardless of where you found a piece of research evidence, you'll need to assess how reliable and relevant it is for your purpose and context before you rely on it to make a decision.

Often, the research evidence you can access will be filtered information that provides analysis, synthesis, interpretation, commentary and/or evaluation of original research studies. In other words, someone else has already interpreted the research for you.

Use the CRAAP test to assess non-academic sources

CRAAP is an acronym for Currency, Relevance, Authority, Accuracy and Purpose. Developed by a librarian at the University of California,[1] it's a set of questions to think about when assessing the extent to which you should rely on a particular non-academic source of evidence such as a website or blog. The questions offer a structured way to decide if the information is likely to be objective and reliable, or whether there are signs it could be irrelevant or biased.

Some questions will be more important than others depending on your purpose, so there are no hard and fast rules. The CRAAP test is a tool to help you – it doesn't replace your professional judgment.

Many of the questions relate to online information, but you can also use the CRAAP test to assess printed texts such as books.

AERO has created a handy template for scoring a source of evidence against the questions. This can help you decide how reliable the source is. It includes some completed examples if you're looking for guidance for scoring.

Is your source CRAAP?

Currency – is the information timely?

- When was the information written or posted? Is it up-to-date, or is there a possibility that the information (or the sources the author draws on) might be out-of-date for the topic? Does information about this topic change rapidly?
- If there are links, are they functional?

Relevance – is the information relevant to your topic?

- Does the information help answer your question or tell you what you need to know?
- Is the information at the right level for you (not too advanced and not too basic)?
- What country is it from? Research generated in contexts that are different to yours can still be helpful, but findings may be less directly applicable compared to research generated in your own context (or in contexts very similar to your own).

Authority – are the authors or publishers credible?

- Who is the author? What are their credentials? Are they qualified to write about this topic? What else do they write about?
- If the author is an organisation, what is the nature of their activity? Can you tell who owns or operates the organisation?

- What can you tell from the URL? For example, is it .com or .com.au, .edu, .org, .gov, or .net?
- Is there contact information? Where are they located?
- If a website, is the layout professional?

Accuracy – is the information likely to be correct?

- Is the information supported by evidence? How credible is the evidence?
- Are references provided? Are they current and academic sources?
- Can you verify the information somewhere else?
- Does the language seem objective and free from emotion and sensationalism?
- Are there any signs of political, personal or other biases?
- Is it well-written with no spelling or grammatical errors?

Purpose – is the information likely to be biased or objective?

- What appears to be the purpose of the information? Is it to inform, entertain, persuade, promote or sell a product or service?
- Who is the intended audience?
- Is it likely the author or organisation has an agenda? For example, a business selling a literacy program might publish a summary of evidence supporting the type of approach their program uses while ignoring evidence that supports alternative approaches.

Not all academic sources are created equal

Even if you're reading filtered information in an academic journal, or information that's published by a reputable research institute, some academic papers are more reliable than others. Filtered academic content includes systematic reviews, meta-analyses, and expert opinion.

Systematic reviews and meta-analyses are considered the highest quality type of evidence. If you find a recent systematic review that answers the question you're interested in, it's usually safe to rely on its findings. That said, just because a paper is labelled as a systematic review, doesn't mean you should automatically trust it. Consider whether it could be biased – for example, in the way it frames the issues being examined or the criteria it uses to select studies for the review. In addition, reviews can only analyse research that has been published.

Systematic reviews aim to provide a robust answer to a particular question by identifying and synthesising all the relevant academic research. They use rigorous and transparent methods to search for and summarise studies. These methods aim to reduce bias and are reported in such a way that another researcher should be able to reproduce the results following the same method. Systematic reviews will also identify when different studies about the same thing have found different results.

Rapid reviews and scoping reviews also use rigorous and transparent methods. Rapid reviews are like systematic reviews with some steps omitted so they can be completed more quickly. This means they may be less comprehensive. Scoping reviews investigate the size and scope of research literature on a topic – for example, to identify gaps in evidence.

Meta-analyses use statistical methods to combine data from multiple studies about the same question to produce a more reliable estimate of the size of the effect of an intervention. They're usually – but not always – based on a systematic review.

Be aware – not all literature reviews are systematic reviews. Some literature reviews simply provide background for a new study or summarise what's known about a topic. They often include little to no details on how they searched the literature, and may be biased if authors highlight studies that support the story they're telling and downplay those that present alternative findings. A systematic review will always be clearly named as such and will include details of the search methods used.

Expert opinion – the opinion of one or more people who are considered to be (or claim to be) experts in the field – is another form of filtered information found in academic journals. Expert opinion based on the experts' knowledge of the research literature may be reliable. Nevertheless, experts sometimes disagree. Perhaps they haven't read all the evidence, or just like everyone else their thinking is unintentionally influenced by bias (see Table 1 in AERO's The Value of Research Evidence practice guide for a list of common cognitive biases). Expert opinion can be a good starting point but rather than taking it at face value, you could ask yourself a few questions:

- Does the expert have an agenda or any conflicts of interest?
- Have they acknowledged potential biases in their work?
- Have they explained the evidence behind their opinion?
- Can you find any other supporting evidence?
- Are there other experts who disagree?

Single studies (unfiltered information) are essential when you want to dive deeply into a topic, when systematic reviews are outdated or simply don't exist, and when you're not sure if an existing review applies to your context (for example, the review didn't include any studies from Australia).

Even when a study is published in a peer-reviewed academic journal, it's important to critically assess what you read, which can be difficult when you're not a trained researcher. Table 1 provides an overview of AERO's evidence use resources to help you reflect on the rigour and relevance of research about a policy, practice or program you're considering implementing.

Table 1: AERO's evidence use resources for assessing rigour and relevance of education evidence

Resource	This resource...	Use it when...
Standards of Evidence	**explains** how much confidence you can have in different types of evidence.	you want to **build your understanding** of what makes rigorous and relevant evidence.
Interactive Evidence Decision-Making Tool	guides you through **2 steps**: 1. assessing your level of confidence in the evidence you currently have for a practice or program 2. deciding what to do next based on your level of confidence.	you want to check the evidence sitting behind a **program or practice** and decide on next steps
Research Reflection Guide	has a set of **questions** to help you consider a piece of research about a practice or program, prompting you to reflect on: • what the research says • how relevant the research is to your context • whether you should implement the approach • what you could do to ensure successful implementation.	you have a piece of **research** (for example, a journal article or research report) about a practice or program you're considering implementing and want to: • assess its rigour and relevance • decide whether to implement the practice or program.
Assessing Whether Evidence is Relevant to Your Context	has a step-by-step **process** for assessing whether a piece of evidence is relevant to your **context**.	you have a piece of **evidence** and want to reflect on whether it's relevant to your context – and therefore, whether the corresponding approach is likely to be effective in your school or service.

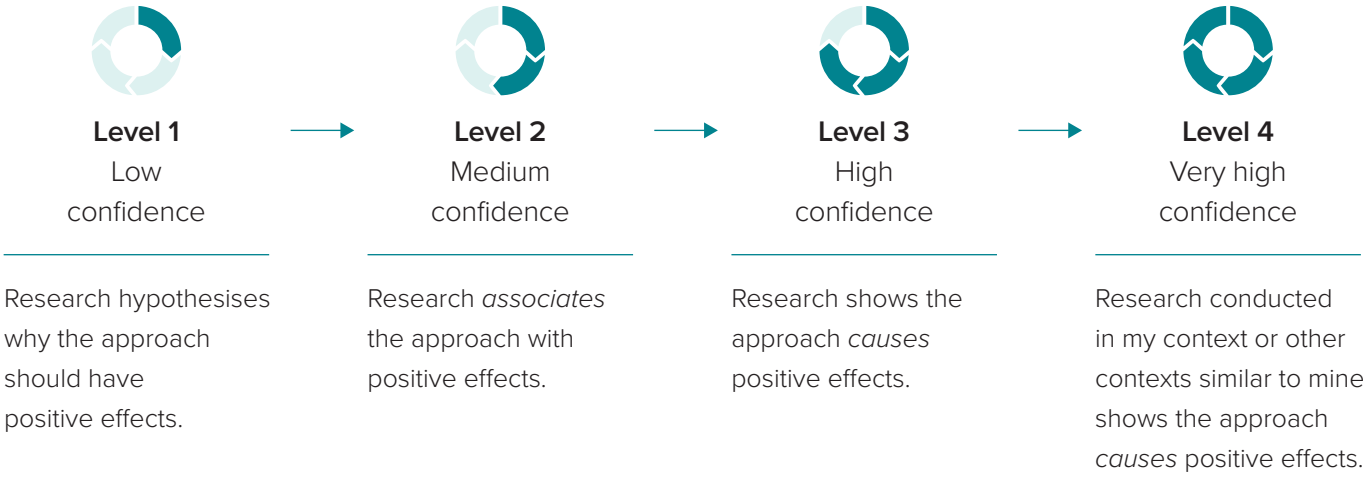
Use AERO's Standards of Evidence to assess rigour and relevance

Standards of evidence help AERO and the education community make consistent and transparent judgements when assessing evidence about the effectiveness of a particular education policy, practice or program.

AERO's Standards of Evidence establish AERO's view on what constitutes **rigorous** and **relevant** evidence about the effectiveness of a particular education approach. 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.

Figure 1: AERO's Standards of Evidence



Rigorous evidence: Cause versus correlation

When assessing evidence about the effectiveness of a particular education policy, practice or program, we need rigorous evidence that isolates the specific impact of the approach we're interested in and rules out the effects of as many other influences as possible. Otherwise, we can't be sure how much of a change was due to the approach and how much was caused by something else or by chance. That's why the AERO Standards of Evidence distinguish between research that *associates* an approach with positive effects (correlational evidence) and research that shows an approach *causes* positive effects.

'Correlation does not imply causation' is a cliché in research, yet it's still common for correlational research findings to be interpreted as if they were evidence of causation. Being able to distinguish between correlational and causal evidence is critical to assessing the strength of evidence supporting an approach.

> When researchers say that 2 variables are linked or that one variable predicts another, they're merely saying that the data are correlated, not that one variable causes the other.
>
> Similarly, an 'effect size' represents the strength of an association, which may or may not be a causal effect.

Correlational evidence

Correlation or association occurs when 2 things tend to move together – that is, when one increases or decreases the other follows a similar pattern. This could mean that one thing is causing the other, but it could also mean that some third variable is influencing both, or it could simply be coincidence.

Correlations may or may not be causal. Causation is when one element, factor or event is known to cause another (for example, a particular teaching practice is known to lead to improvements in student test scores).

Other correlations may seem causal, but research shows us that this is not the case. Evidence for Learning provides an example of this in Top Ten Education Myths About What Improves Student Learning when they

seek to dispel the myth that introducing school uniforms supports a culture of discipline, which in turn causes an improvement in student achievement. The article notes:

> [T]here is no robust evidence that introducing a school uniform will, by itself, improve performance, behaviour or attendance. There are studies about these outcomes linked to the introduction of a school uniform policy, but uniform was usually one factor amongst other improvement measures, such as changes in behaviour policy or other teaching and learning developments. So, whilst there is a link between schools with good academic outcomes and uniforms, there is no evidence that uniforms are the cause.

Also, relationships between variables can be complex. In the 1980s, research reporting a correlation between self-esteem and success led to the development of numerous programs to increase self-esteem as a means to increase school achievement. Later research showed that the relationship is reciprocal (success causes high self-esteem at least as much as high self-esteem causes success), that both are influenced by other factors such as home environment or stressful life events, and that different facets of self-esteem might operate differently.[2]

Causal evidence

To establish that one thing (let's call it X) causes another (let's call it Y), research needs to show 3 things:

- X comes before Y in time
- the relationship isn't caused by chance
- the relationship isn't caused by other factors.

When it comes to finding out whether certain programs or practices are effective, we want to know if the program or practice causes a change in the outcome we're aiming for, not just whether the practice and the outcomes are correlated. The best way to do this is through a well-designed randomised controlled trial (RCT).

In an RCT, participants are randomly assigned to one of two groups: one receiving the approach that is being tested (the experimental or treatment group), and the other receiving an alternative approach or no approach (the comparison or control group). That is, which group an individual is in is determined by chance alone. This helps ensure that the 2 groups are equivalent at the start and any differences in results are due to the program, not to chance or other factors.

Why random assignment matters

Imagine we're designing an experiment to test whether a new reading program is more effective than the existing program. If each teacher chooses whether to use the new program, teachers who are more confident might choose the new program while those who are less confident or have more challenging students might choose to stick with what they know. If we find that students who did the program have better results, we can't be sure that it wasn't due to the confidence of their teacher or the nature of the students. Random assignment means that these kinds of *confounding factors* should be spread equally across both groups.

Experiments with people in the real world are never perfect because it's impossible to control everything. If teachers are delivering a program, it can be very difficult to make sure they all deliver it exactly the same way, and simply knowing you're in an experiment and being observed can change your behaviour. Nevertheless, when you're reading research that aimed to test whether one approach was more effective than another, an RCT provides the most reliable evidence.

Of course, not all high-quality research is an RCT. RCTs aren't always possible or appropriate. They can be expensive to run, take a long time, require a large number of participants, and aren't suited to all research questions. For example, if researchers want to study whether boredom during lessons causes low school attendance, they can't randomly assign students to be bored during lessons. But if researchers want to study whether a new small group program for reducing boredom causes higher school attendance, they could look at students who already report being bored in lessons. They could randomly assign some of these students to receive the new program and others to continue in their usual lessons, and answer this question with an RCT.

Be aware – when it comes to specific populations, there might not be much causal evidence available. For example, much of the education research conducted with First Nations communities or people living in very remote areas is descriptive or correlational. It's often impractical to conduct RCTs with small population groups because they require large numbers of participants. Researchers also carefully consider the ethical implications of different research designs, such as the appropriateness of control groups, addressing informed consent and maintaining the privacy and confidentiality of research participants. These and other ethical issues (for example, conducting research in a culturally responsive way) are particularly important for research involving equity groups. For more information, see the Additional considerations when assessing evidence section of this practice guide.

Respectful First Nations education research must include and privilege First Nations stakeholders at every stage, ensuring that their outcome priorities drive the research agenda. The strength of research conducted in this way will result in deep and rich data that will provide a holistic understanding of the many interrelated factors that impact learning outcomes for First Nations children and young people.

It's important that any research undertaken with and for First Nations children, families and communities should:

- privilege their voices and implement an Indigenist way of working that prioritises how First Nations Peoples want research to be undertaken, recorded, accessed, reported on, shared and stored
- has merit and integrity
- prioritise the development of relationships of trust and reciprocity between researchers and First Nations communities
- provide clear benefit to First Nations communities, families and children
- be undertaken by researchers who are culturally responsive and use reflexive practice to improve their knowledge and understanding.

Quasi-experimental designs are often used when random assignment isn't possible. For example, if a few teachers have decided to try the new reading program and others are sticking with the existing program, researchers could compare these pre-existing groups. However, the extent to which you can be sure of a cause and effect relationship will depend on how well the researchers are able to account for confounding variables. Confounding variables are other factors besides the variable of interest that could influence outcomes. In this example, the reading program is the variable of interest. The teacher is a confounding variable because teachers who decided to use the new program may have different characteristics from those using the old program. We can't rule out the possibility that any difference in outcomes will be the result of the teachers in one group being different from the other group, rather than a result of the program. In quasi-experimental designs, complex statistical methods can be used to help control for differences between the groups. However, it's almost impossible to account for every confounding variable.

Longitudinal studies are another common design that measures change over time. In the case of the reading program, researchers might measure students' reading skills multiple times before, during and after the reading program. If the data showed that students' scores increased more rapidly while they were doing the program, it might seem reasonable to conclude that the program was effective. But without a control group and random assignment, it's possible that a third variable caused the change. Perhaps introducing a new program energised the teacher and that contributed to better learning. Perhaps a different teacher taught the new program and it was actually the change in teacher that improved learning. Longitudinal designs are valuable for finding out how things change over time but are not a replacement for experimental designs.

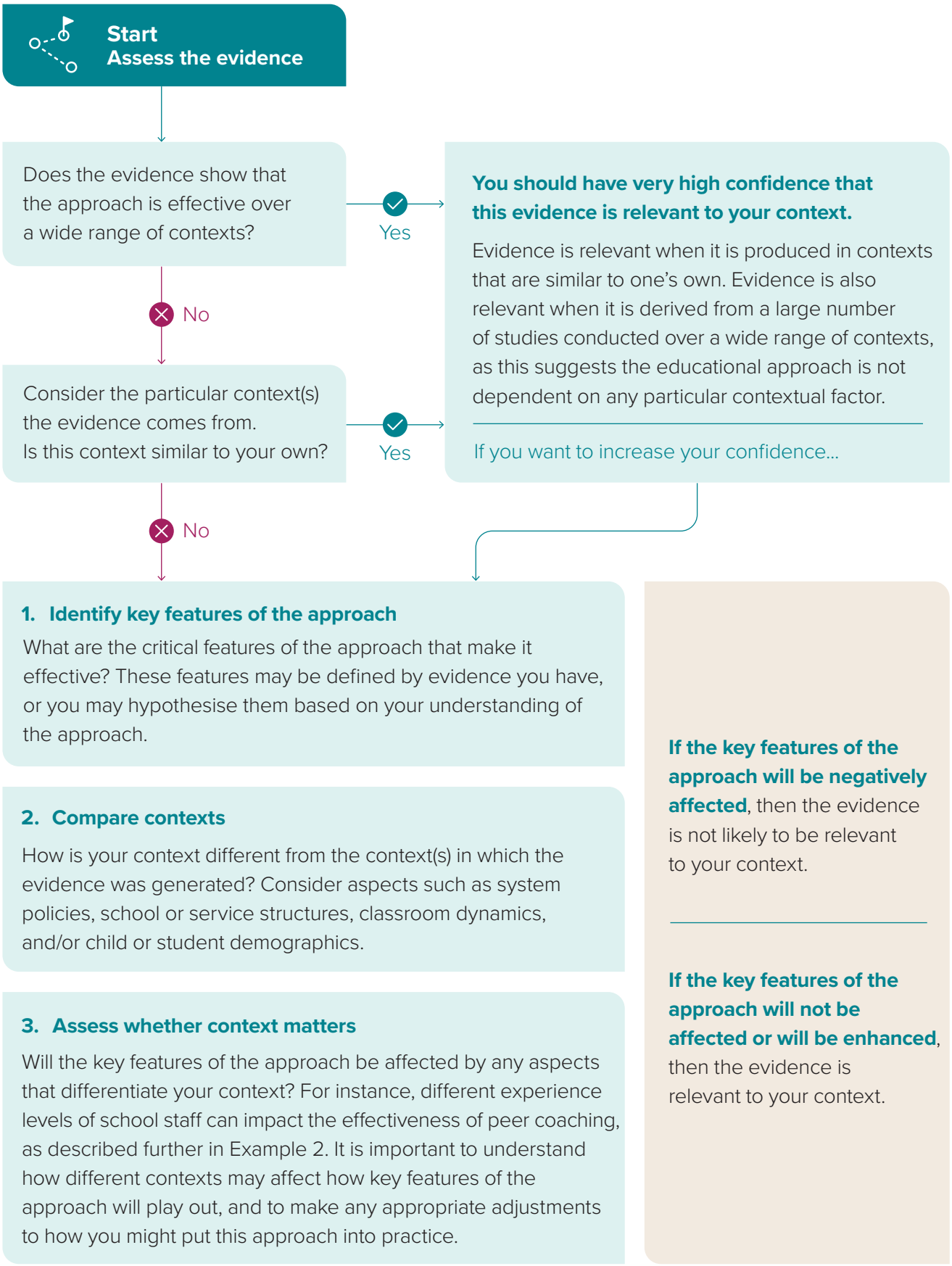
Relevant evidence: Will this work in my context?

It's common to wonder whether evidence from research or from other educators or teachers applies to your context. Perhaps the evidence is from a different type of community (regional or remote versus metropolitan), a different type of school (primary versus secondary) or service (family day care versus long day care centre) or with different children (high versus low socio-economic status). Given these differences, how do you know whether a practice or program supported by research will also be effective in your school or service?

Concerns about context shouldn't stop you from engaging with research that has strong potential to transform your practice.

AERO's Assessing Whether Evidence is Relevant to Your Context practice resource can help you reflect on whether a piece of evidence is relevant to your context – and therefore, whether the corresponding approach is likely to be effective in your school or service. The resource takes you through a step-by-step process (Figure 2) for assessing the relevance of evidence.

Figure 2: Process for assessing relevance



Additional considerations when assessing evidence

This practice guide and AERO's Standards of Evidence prioritise 2 criteria for assessing education evidence: **rigour** and **relevance**. These criteria have been prioritised because they're 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.

When assessing research, keep in mind that it should be meaningful and valuable for the communities involved. 'Respect and recognition' is one of the principles for the ethical and responsible conduct of research underpinning AERO's research. This involves meaningful engagement with communities to understand their context and perspectives.

The article featured in AERO's Research Reflection Guide worked example illustrates the importance of developing research that is culturally responsive. The research, conducted with Aboriginal children in the Northern Territory, found that an early childhood playgroup program that embedded culture and local language was associated with positive early learning and language outcomes. Examples from this research of practices aligned with a culturally responsive approach include:

- collaboration with communities to ensure program integrity and cultural appropriateness
- local Aboriginal staff delivering the program in first language and providing support for culturally appropriate adaptations
- adapting an early childhood screening assessment so it was a culturally and linguistically relevant tool to assess Aboriginal children's outcomes.

The authors note that ownership by staff and community members was key to the success of the program.

For more information about the ethical practice of research concerning First Nations people, see the AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research.

Knowledge check: Assessing research evidence

Take our quick quiz below or scan the QR code to test your knowledge about research evidence.

1. If a paper has been published in a peer-reviewed academic journal, you don't need to assess the rigour and relevance of the study.

a. True

b. False

c. It depends on the journal

2. Expert opinion...

a. should be trusted because the experts have analysed the research and come to a consensus

b. can be subjective because, just like everyone else, experts can be unintentionally influenced by bias

c. is as reliable as a systematic review if it's published in a peer-reviewed journal

3. A researcher collected data about use of a new reading program and student test scores to find out if the new program was more effective than the existing program. Which of the following statements is TRUE? The research would be considered rigorous if:

a. The research methods ruled out the effects of as many other influences on test scores as possible.

b. The research found that children in the new program achieved higher test scores than children in the existing program.

c. Teachers reported that children enjoyed the new program more than the existing program.

4. According to AERO's Standards of Evidence, you should have very high confidence in an approach if you have research that associates the approach with positive effects.

a. True

b. False

5. To be relevant, a study MUST be conducted:

a. with students or young people who are the same age as those in my service or school

b. in Australia, and preferably in my state

c. using a rigorous experimental design

d. in similar settings to my own (though what's 'similar' will depend on what's important for the topic)

6. You come across 4 studies talking about positive effects of a program on infants' adjustment to early education and care. Which of the studies appears to have evidence of causation?

a. ... *leads to* positive effects ...

b. ... is *linked to* positive effects ...

c. ... *predicts* positive effects ...

d. ... is *associated with* positive effects ...

7. A systematic review:

a. uses statistical methods to combine data from multiple studies to produce a more reliable estimate of the size of the effect of a program

b. clearly and methodically reports on the results of a single trial testing the effect of a program

c. uses rigorous methods for searching, synthesising and reporting on multiple studies about the same question

8. A researcher observes that 2 variables, A and B, move together. That is, when A increases, B also increases. Which of these statements about A and B is TRUE?

a. Variable A and B are correlated

b. Variable A caused variable B to increase

c. Variable B caused variable A to increase

Next steps

Once you've determined the evidence is relevant, you can use AERO's Interactive Evidence Decision-Making Tool to help determine how to implement the approach.

For an overview of different types of evidence and biases to be aware of when reading about evidence, see AERO's The Value of Research Evidence practice guide.

For guidance on how to find research evidence, see AERO's Looking for Research Evidence practice guide practice guide.

For guidance on how to apply research evidence, see AERO's Applying Research Evidence practice guide.

Answers: 1(b), 2(b), 3(a), 4(b), 5(d), 6(a), 7(c), 8(a).

Endnotes

1 Blakeslee, S. (2004). The CRAAP Test. *LOEX Quarterly*, *31*(3), Article 4. https://commons.emich.edu/loexquarterly/vol31/iss3/4/

2 See:
Baumeister, R. F., Cambell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest*, *4*(1). https://doi.org/10.1111/1529-1006.01431
Kärchner, H., Schöne, C., & Schwinger, M. (2021). Beyond level of self-esteem. *Social Psychology of Education*, *24*, 319–341. https://doi.org/10.1007/s11218-021-09610-5
Midgett, J., Ryan, B. A., Adams, Gerald R., & Corville-Smith, J. (2002). Complicating achievement and self-esteem: Considering the joint effects of child characteristics and parent–child interactions. *Contemporary Educational Psychology*, *27*(1), 132–143. https://doi.org/10.1006/ceps.2001.1083