

Standards of Evidence:

A common language for decision making, implementation and evaluation in schools

Matt Deeble

Director

Social Ventures Australia

Dr Deb Kember, Deb Fisher,

State Schools Division

Department of Education, Queensland

While you are waiting.....

To receive session resources in a digital pack:



- add your email address to one of the recording sheets
- resources will be emailed to you before the end of the conference.

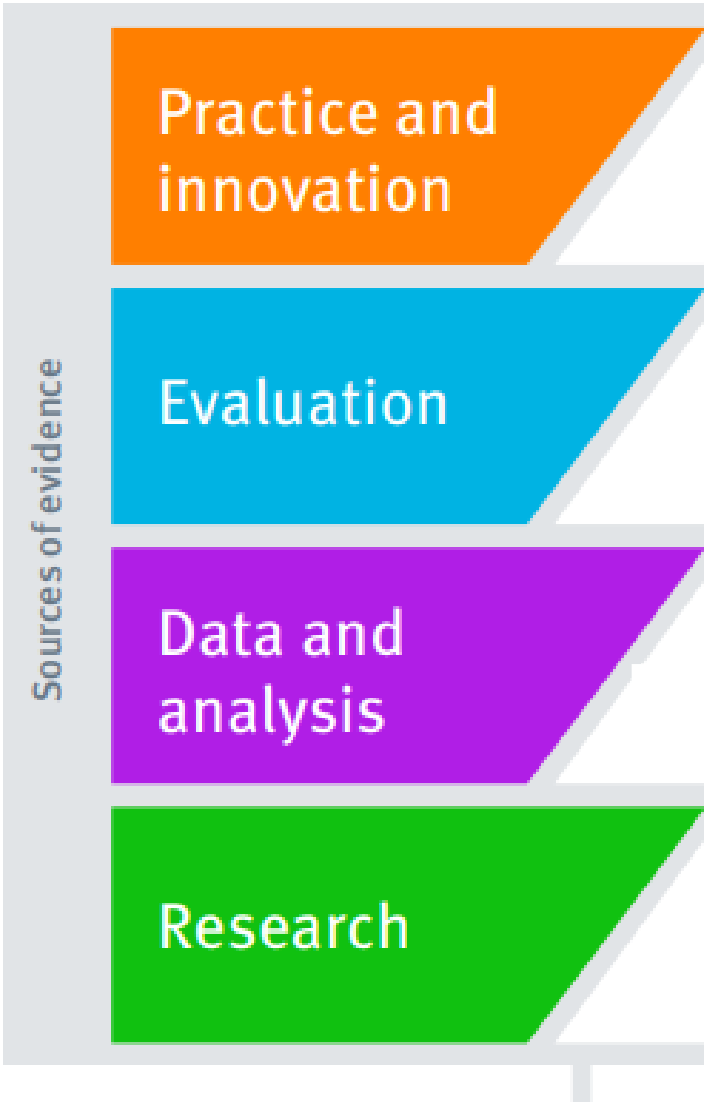


Queensland's state schools

- 552,512 fulltime students
- 1240 state schools
- 7 educational regions
- Queensland is home to the fastest improving educational system in Australia



Evidence Framework



- 🎯 can generate reliable evidence of what works
- 🎯 plays in the design and implementation of the department's initiatives
- 🎯 quality, robust data informs our initiatives and helps us measure our impact
- 🎯 rigorous research helps ensure our work is informed by the best available evidence



The Standards of Evidence

Design

Quality of attribution

Impact

Measurable change

Scalability

Potential to implement

Investment

Creating value

Strength of evidence	<p>Verified attribution</p> <p>An independent source or sources verifies the demonstrated measurable change can be attributed to the initiative</p>	<p>Very high</p> <p>Very large measured improvement</p>	<p>Larger scale implementation</p> <p>An independent source or sources verifies the initiative has been implemented by others on a larger scale</p>	<p>Cost-effectiveness known</p> <p>Information on cost-effectiveness exists</p>
	<p>Attribution</p> <p>Evidence from a high-quality study or studies establishes the measurable change can be attributed to the initiative</p>	<p>High</p> <p>Large measured improvement</p>	<p>Implemented elsewhere</p> <p>The initiative can be implemented in more than one setting and/or with different groups with at least the same impact. Documentation supports implementation</p>	<p>Outcomes costed</p> <p>Information on the costs per unit of outcome exists</p>
	<p>Possible attribution</p> <p>A reasonable explanation supports attribution of the measurable change to the initiative</p>	<p>Moderate</p> <p>Medium measured improvement</p>	<p>Well documented</p> <p>Documentation exists about how the initiative is implemented and is comprehensive enough to enable implementation with different groups or in other settings</p>	<p>Outputs costed</p> <p>Information on the costs per unit of output exists</p>
	<p>Attribution unclear</p> <p>A logical plan describes the initiative and its objectives, target and comparison groups, inputs, outputs, outcomes and timelines. It includes the measures and methods for collecting, storing and analysing data</p>	<p>Low</p> <p>Small measured improvement that can be reasonably linked to the initiative</p>	<p>Could work</p> <p>Documentation describing the implementation of the initiative in one setting exists</p>	<p>Costs known</p> <p>The costs of the initiative (for example, financial, human resources, infrastructure, program licensing, implementation etc.) are known</p>
	<p>No attribution possible</p> <p>The plan or design of the study and the quality of the data or monitoring are not sufficiently robust to support attribution</p>	<p>Unknown</p> <p>Impact cannot be measured or unintended impact is identified</p>	<p>Unknown</p> <p>Insufficient documentation exists on the initiative</p>	<p>Unknown</p> <p>There is no or limited information on the cost</p>

Our common language

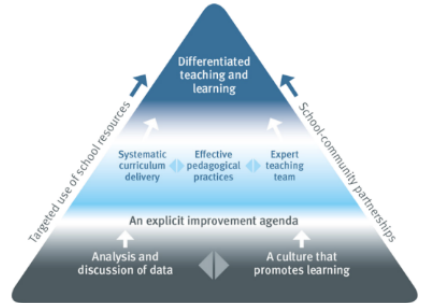
State Schools Strategy Our strategic priorities



School improvement model

School Improvement Hierarchy

Where we are



Inquiry cycle

How we learn



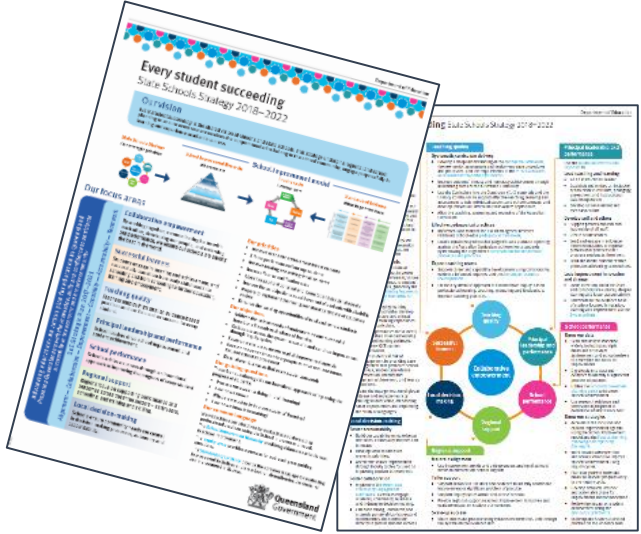
Standards of Evidence

What impact we have

	Design Quality of attribution	Impact Measurable change	Scalability Potential to implement	Investment Creating value
High	High An independent reviewer or review panel, the observational researcher change can be attributed to the initiative	Very large measured improvement	Large number of schools An initiative can be implemented in more than one setting or across different groups with at least the same impact. Dissemination requires implementation	Extensive impact Information on the costs per unit of success rates
Medium	Medium Evidence from a high quality audit or studies, evaluation of the researcher change can be attributed to the initiative	Large measured improvement	Medium number of schools The initiative can be implemented in more than one setting or across different groups with at least the same impact. Dissemination requires implementation	Substantial impact Information on the costs per unit of success rates
Low	Low A researcher or reviewer supports attribution of the researcher change in the initiative	Medium measured improvement	Small number of schools Dissemination can be implemented in more than one setting or across different groups with at least the same impact. Dissemination requires implementation	Some impact Information on the costs per unit of success rates
Very low	Very low A high level of evidence of the initiative and its objectives, impact and implementation quality, including the researcher and evidence for collecting, timing and implementation	Small measured improvement that can be reasonably linked to the initiative	Small number of schools Dissemination can be implemented in more than one setting or across different groups with at least the same impact. Dissemination requires implementation	Minimal impact Information on the costs per unit of success rates
Not applicable	Not applicable The plan or design of the study and the quality of the data are insufficient to support attribution	Impact cannot be measured or attributed to the initiative	Insufficient dissemination across the initiative	Insufficient There is no or limited information on the cost

Guiding questions

1. How are our students doing in their learning?
2. How do we know?
3. What are we doing to improve students' learning?
4. How do we know it is working?

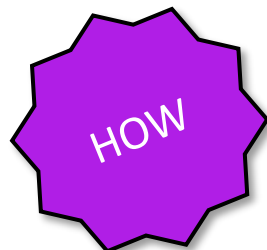


Queensland state schools improvement agenda



Every student succeeding

- ✓ Engage in inquiry
- ✓ Apply Standards of Evidence
- ✓ Access local evidence
- ✓ Draw on research
- ✓ Discuss evidence



Evidence-informed continuous improvement



Dimensions in detail



The **Design** dimension describes the extent to which any measurable change can be attributed to an initiative.



The **Impact** dimension is concerned with the strength of measurable change.



The **Scalability** dimension indicates the extent to which an initiative could be/has been reproduced with similar or greater success across other settings or groups.



The **Investment** dimension is concerned with return on investment based on the evidence of cost-effectiveness or value for money.

The Standards of Evidence

The image displays four overlapping documents, each representing a different calibration process under the 'Standards of Evidence' framework. The documents are arranged from top to bottom, with the 'Design calibration process' document being the most prominent and detailed.

- Standards of Evidence Investment calibration process:** Features a green dollar sign icon.
- Standards of Evidence Scalability calibration process:** Features a purple bar chart icon.
- Standards of Evidence Impact calibration process:** Features an orange flame icon.
- Standards of Evidence Design calibration process:** Features a blue triangle icon and a detailed flowchart.

The 'Design calibration process' flowchart includes the following questions and decision paths:

- Question 1: "Is there an initiative that a measurable change in performance is connected with a specific initiative?"
 - Yes: Proceeds to Question 2.
 - No: Proceeds to the 'Not a Design Initiative' box.
- Question 2: "Are the initiative's objectives and measures identified in the study design?"
 - Yes: Proceeds to Question 3.
 - No: Proceeds to the 'Not a Design Initiative' box.
- Question 3: "Does the study design support the possible attribution of the results to the initiative?"
 - Yes: Proceeds to Question 4.
 - No: Proceeds to the 'Not a Design Initiative' box.
- Question 4: "Does the high quality study design establish attribution of the results to the initiative?"
 - Yes: Proceeds to the 'High Quality Attribution' box.
 - No: Proceeds to Question 5.
- Question 5: "Has the study design been verified to high quality by an independent source?"
 - Yes: Proceeds to the 'High Quality Attribution' box.
 - No: Proceeds to the 'Not a Design Initiative' box.
- Question 6: "Is there a change in performance of the identified measures?"
 - Yes: Proceeds to the 'High Quality Attribution' box.
 - No: Proceeds to the 'Not a Design Initiative' box.
- Question 7: "Has the initiative been implemented?"
 - Yes: Proceeds to the 'High Quality Attribution' box.
 - No: Proceeds to the 'Not a Design Initiative' box.

At the bottom of the flowchart, there are six boxes representing the final classification of the initiative:

- Not a Design Initiative (The DET tool design dimensions does not apply to the type of initiative)
- Low Quality Attribution
- High Quality Attribution
- High Quality Attribution
- High Quality Attribution
- High Quality Attribution





Quality of attribution

Verified attribution

An independent source or sources verifies the demonstrated measurable change can be attributed to the initiative

Attribution

Evidence from a high-quality study or studies establishes the measurable change can be attributed to the initiative

Possible attribution

A reasonable explanation supports attribution of the measurable change to the initiative

Attribution unclear

A logical plan describes the initiative and its objectives, target and comparison groups, inputs, outputs, outcomes and timelines. It includes the measures and methods for collecting, storing and analysing data

No attribution possible

The plan or design of the study and the quality of the data or monitoring are not sufficiently robust to support attribution

- What was the research question/focus? How is this relevant to our priority or problem of practice (and context)?
- (Precisely) How was it implemented? Who was involved and what were their attributes? What were the inputs – structures, processes, resources?
- What were the measures of success – implementation and outcomes? How did they monitor implementation and change?
- Was the methodology robust? (sample size, comparison data, analyses) How was research selected/rejected for meta analysis?



Measurable change

Very high

Very large measured improvement

High

Large measured improvement

Moderate

Medium measured improvement

Low

Small measured improvement that can be reasonably linked to the initiative

Unknown

Impact cannot be measured or unintended impact is identified

Did the research/review/meta analysis report a measurable difference to (targeted) student outcomes?

Did they supply baseline and endpoint data? Can we check their analyses, interpretations and conclusions?

How did they describe their claims of strength of impact?

What assumptions and predictions (bias) can we detect? Can we accept them?

How will we confirm/triangulate across multiple measures?

How will we determine the strength of impact?



Scalability

Potential to implement

Larger scale implementation

An independent source or sources verifies the initiative has been implemented by others on a larger scale

Implemented elsewhere

The initiative can be implemented in more than one setting and/or with different groups with at least the same impact. Documentation supports implementation

Well documented

Documentation exists about how the initiative is implemented and is comprehensive enough to enable implementation with different groups or in other settings

Could work

Documentation describing the implementation of the initiative in one setting exists

Unknown

Insufficient documentation exists on the initiative

- Was this conducted in one or more settings?
- How much information is available? Enough to implement (with ‘fidelity’)?
- Was there any independent verification of implementation at scale, with impact?

Investment

Creating value

Cost-effectiveness known

Information on cost-effectiveness exists

Outcomes costed

Information on the costs per unit of outcome exists

Outputs costed

Information on the costs per unit of output exists

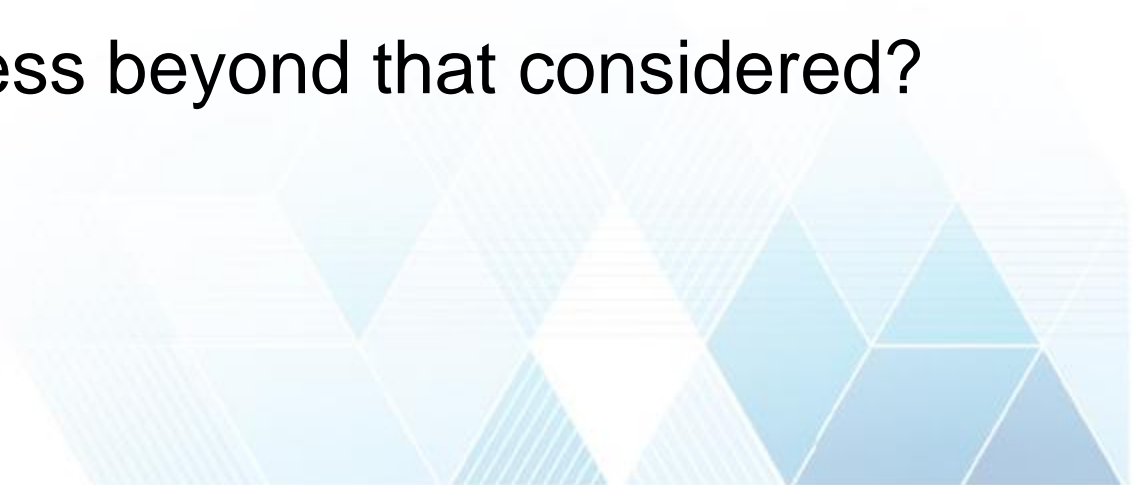
Costs known

The costs of the initiative (for example, financial, human resources, infrastructure, program licensing, implementation etc.) are known

Unknown

There is no or limited information on the cost

- Were the resources described/listed/detailed?
- Were they costed?
- Was cost per student determined?
- Was cost-effectiveness beyond that considered?





Quality of attribution

Verified attribution

An independent source or sources verifies the demonstrated measurable change can be attributed to the initiative

Attribution

Evidence from a high-quality study or studies establishes the measurable change can be attributed to the initiative

Possible attribution

A reasonable explanation supports attribution of the measurable change to the initiative

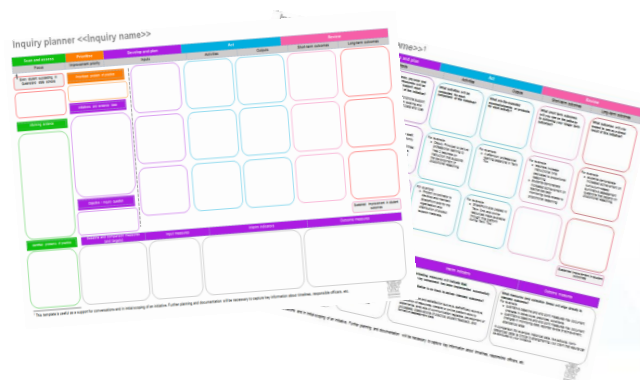
Attribution unclear

A logical plan describes the initiative and its objectives, target and comparison groups, inputs, outputs, outcomes and timelines. It includes the measures and methods for collecting, storing and analysing data

No attribution possible

The plan or design of the study and the quality of the data or monitoring are not sufficiently robust to support attribution

- What is the **problem of practice** being addressed? What was the **evidence** that led this being prioritised?
- What is your **improvement initiative**? What is its objective? How will be implemented? Who and what will be involved?
- What are the **measures of success**? How will the initiative be **monitored** to gauge progress? What were the interim measures? Was there a logical chain of evidence?
- Is there **alignment** between the problem of practice, the improvement initiative and the measures of success?



**Very high**

Very large measured improvement

High

Large measured improvement

Moderate

Medium measured improvement

Low

Small measured improvement that can be reasonably linked to the initiative

Unknown

Impact cannot be measured or unintended impact is identified

How will we know if the initiative making a difference to **student outcomes**?

Do we have **baseline** and **endpoint data**? How will we **analyse** (and interpret and represent) data?

What assumptions and predictions (bias) do we have?

How will we confirm/triangulate across **multiple measures**?

How will we determine the **strength** of impact?





Scalability

Potential to implement

Larger scale implementation

An independent source or sources verifies the initiative has been implemented by others on a larger scale

Implemented elsewhere

The initiative can be implemented in more than one setting and/or with different groups with at least the same impact. Documentation supports implementation

Well documented

Documentation exists about how the initiative is implemented and is comprehensive enough to enable implementation with different groups or in other settings

Could work

Documentation describing the implementation of the initiative in one setting exists

Unknown

Insufficient documentation exists on the initiative

- What will we need to do/document to ensure sustainability, allow scaling (or inform adjustments)?
- What could we consider – RCTs, clusters, partners, external evaluators?
- How do we verify? How do we share?
- How do we support others to ‘replicate’ in context?

Investment

Creating value

Cost-effectiveness known

Information on cost-effectiveness exists

Outcomes costed

Information on the costs per unit of outcome exists

Outputs costed

Information on the costs per unit of output exists

Costs known


The costs of the initiative (for example, financial, human resources, infrastructure, program licensing, implementation etc.) are known

Unknown

There is no or limited information on the cost

- What **resources** (HR, financial, IT, other?) will be required to implement the improvement initiative?
- Is there an understanding of the **total cost** of this improvement initiative?
- How will we keep track of the costs directly related to this initiative?
- How could we calculate the cost of this improvement initiative **per student or teacher** involved?
- What more would we need to do to determine the cost per outcome?

Our invitation to you

- Positive, potential, promise
 - Concerns, constraints, cautions
 - Interesting, blind spots, improvements, applications
- 

Documenting our learning

Inquiry planners



Actionable playbooks

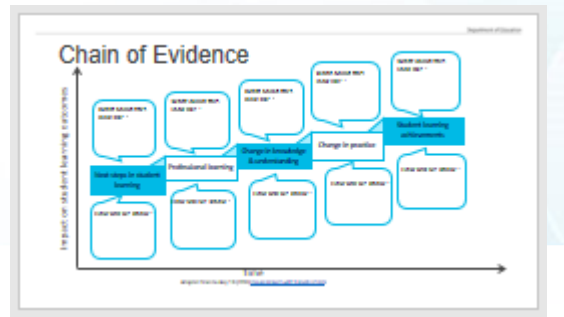
Conversation & calibration tools

The SIM GPS



Generated local evidence

Evidence chains



Dr Deb Kember, Director, School Improvement

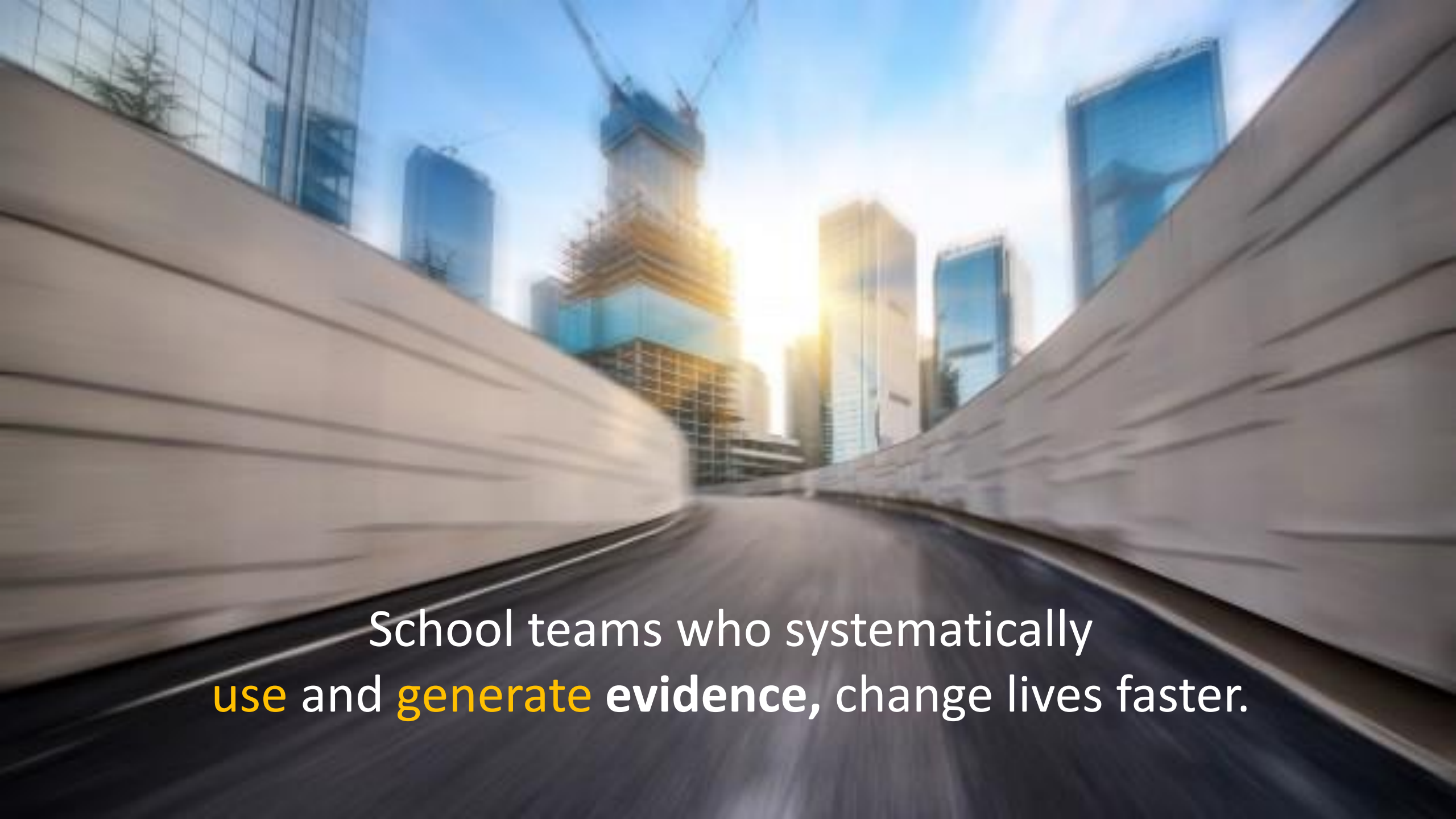
Deb.kember@qed.qld.gov.au

Deb Fisher Director School Improvement

Deb.fisher@qed.qld.gov.au

State Schools Division

Department of Education, Queensland

A blurred city street at sunset with a construction site in the background. The sun is low on the horizon, creating a warm glow and lens flare. The buildings are modern and glass-fronted. The street is flanked by concrete walls, and the overall scene is in motion, suggesting speed and progress.

School teams who systematically
use and **generate** evidence, change lives faster.