Sax Institute

WORKSHOP · 60 MINUTES

Process tracing in practice: testing causal claims with mixed evidence

Dr Kizzy Gandy, Head of Evidence and Impact, Sax Institute

Thursday 17 September 2026 · 1:30–2:30pm ACST · Rooms 3+4

WHY PARTNER WITH SAX INSTITUTE

The Sax Institute is Australia's most trusted partner for connecting evidence and decisions

Independent. Not-for-profit. Putting evidence to work for over 20 years.

OUR VISION

A healthier future for every generation.

WHAT WE BRING

Australia's broadest range of evidence tools and methods, and an unrivalled partner network.

HOW WE HELP

Independent, end-to-end intelligence that improves health and wellbeing systems.

A NOT-FOR-PROFIT THAT CONNECTS EVIDENCE AND DECISION-MAKING

Assets and capabilities no other Australian organisation can match

45 and Up & 18 and Up Studies

250,000+ participants over 20 years. Quality data on ageing, chronic disease and mental health.

SURE

Australia's most trusted secure research platform, serving 60+ data custodians.

Evidence & Impact Services

Modelling, evaluation, training and evidence synthesis to act on complex problems.

80+ member network

Access to Australia's leading research institutions, with rigour general consultancies can't match.

The Australian Prevention Partnership Centre

The only national collaboration dedicated to strengthening the prevention system.

One partner, end to end

From the first question to measured impact, we bring it all together.

Systems intelligence for health and wellbeing

We help organisations see how the parts of a health system interact — so decisions land where they make the greatest difference.

HOW WE WORK WITH YOU

1 Understand

What is really driving the problem?

- Evidence synthesis
- Benchmarking
- System mapping

2 Plan

Which option gives the greatest return?

- Dynamic simulation modelling
- Decision labs
- Strategy & capability

3 Implement

Is delivery on track, and adapting?

- Monitoring frameworks
- Adaptive management
- Behaviour change

4 Evaluate

What impact did it have, system-wide?

- Systems evaluation
- Economic analysis
- Scaling plan

A SYSTEMS LENS LETS YOU

01 **Invest where returns are highest**

02 **Avoid unintended consequences**

03 **Show greater value**

LEARNING OBJECTIVE

By the end of this hour you will be able to construct core and alternative hypotheses from a theory of change, design evidence tests that distinguish between them, and make a defensible causal claim.

EXERCISE 1

Build the hypotheses

EXERCISE 2

Design the evidence tests

EXERCISE 3

Weigh mixed evidence

SESSION PLAN

How the hour runs

TIME	WHAT WE DO	FORMAT
0–05	Why contribution analysis needs a next step	Presentation
05–20	The Embrace evaluation case study	Presentation
20–30	Exercise 1 · Core and alternative hypotheses	Solo, then plenary
30–40	Exercise 2 · Evidence tests	Solo, then plenary
40–52	Exercise 3 · Weighing mixed evidence	Solo, then plenary
52–60	Applying this to your own evaluations	Discussion

PART ONE

Why contribution analysis needs a next step

METHOD

What is theory-based evaluation?

EXPERIMENTAL EVALUATION

Can demonstrate the program **caused** the observed outcomes

THEORY-BASED EVALUATION

Can demonstrate the program **contributed to** observed outcomes (and explain how and why)

OBSERVATIONAL EVALUATION

Can demonstrate an **association** between the program and observed outcomes

METHOD

Mechanisms of change

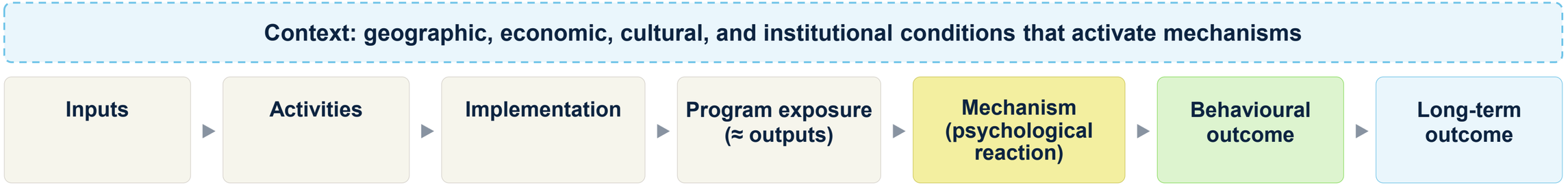
A key feature of theory-based evaluation is examining mechanisms of change which are the **cognitive and emotional reactions** people have to the program. Mechanisms drive outcomes and therefore provide explanatory evidence about how and why the program works.

LOGIC MODEL



No mechanisms. No context. Outputs are assumed to produce outcomes.

THEORY OF CHANGE



Context determines which mechanisms are activated. Mechanisms lead to behaviour change, which produces outcomes.

METHOD

How the data are analysed

Theory-Based Evaluation can be used with all types of data collection and analysis, including surveys, interviews, focus groups, administrative data, ethnographic observations, case studies, and document reviews. Data can be analysed in many ways, including:

Framework Analysis	Looks for indicators of the ToC Change components in qualitative data (pre-specified themes) rather than letting themes emerge from the data, to systematically reduce the data for analysis. This helps to validate the theory.
Descriptive statistics	Summarises the percentage of the sample that achieved the ToC components in quantitative data. This helps to validate the theory.
Contribution Analysis	Looks for qualitative or quantitative evidence to support pathways of change in the ToC model (including mechanisms of change) and considers the influence of other factors on the outcomes. This allows claims that the program contributed to the observed outcomes.
Quasi-experimental study	Tests cause-and-effect hypotheses in the ToC by comparing outcomes between an intervention group and a comparison group that weren't randomised. This further raises the standard of causal evidence within a theory-based approach.
Process Tracing	Tests theorised and counterfactual pathways of change in a ToC using contribution analysis or a quasi-experimental study. This **reduces evaluator confirmation bias** (i.e., only looking for confirmatory evidence) which is a risk associated with observational data.

THE PROBLEM

Where contribution analysis stops

CONTRIBUTION ANALYSIS

Looks for evidence that the theorised pathway of change occurred.

Other influences are considered and discussed — but they are not formally tested. The conclusion is a plausibility judgement.

PROCESS TRACING

Also states what else could have produced the outcome — and tests that.

Core and alternative hypotheses are specified before data collection, each with its own evidence. The conclusion is a comparison.

Process tracing raises the rigour of a causal claim not by collecting more data, but by being deliberate about what the data needs to show — and what it would need to show if the alternative explanation were true instead.

THE METHOD

What is process tracing?

1. **It looks inside a case, not across cases.** Rather than comparing many cases statistically, process tracing follows the chain of events in one case to see whether a hypothesised cause actually produced the outcome through an identifiable mechanism.

2. **It tests for evidence of a causal mechanism.** You specify what “fingerprints” a proposed explanation should have left if it were true, then check the historical record, documents, interviews, and other sources for those specific traces.

3. **It can confirm, disconfirm, or rule out rival explanations.** A strong analysis doesn’t just support the favoured explanation; it actively checks whether competing accounts fit the evidence and eliminates those that don’t.

It uses four evidentiary tests to weigh findings. Straw-in-the-wind (weak, suggestive), hoop (necessary but not sufficient to confirm), smoking gun (sufficient to confirm but not necessary), and doubly decisive (both confirms one explanation and rules out others).

4. **Its strength depends on the quality and specificity of predictions.** Vague expectations produce weak inferences. Rigorous process tracing sets out precise, observable implications in advance, which makes the presence or absence of evidence genuinely informative.

THE METHOD

The four evidentiary tests

FLAG Distinguish between evidence that is unavailable and evidence that is contrary to a hypothesis

After data collection, list each hypothesis (core and alternative) in the relevant quadrant to see which one(s) have the highest standard of evidence.

	WEAK CONFIRMATORY EVIDENCE	STRONG CONFIRMATORY EVIDENCE
WEAK DISCONFIRMATORY	**1 Straw-in-the-wind** All confirmatory and disconfirmatory evidence is weak so no claims about causal inference can be made.	**3 Smoking gun** Strong confirmatory evidence for a hypothesis but only weak disconfirmatory evidence for rival hypotheses so they can't be eliminated.
STRONG DISCONFIRMATORY	**2 Hoop test** DO THIS TEST FIRST If the **necessary** confirmatory evidence is not observed for a hypothesis, it can be eliminated.	**4 Doubly decisive** VERY RARE Strong confirmatory for a hypothesis and strong disconfirmatory for rival hypotheses.

THE CORE QUESTION

“What would we expect to see in the data if this explanation were true — and what would we expect to see if the rival explanation were true instead?”

If both explanations predict the same evidence, that evidence cannot help you. The work is finding the evidence that separates them.

THE METHOD

Process tracing in three steps

1

State the hypotheses

Read the core claim off the theory of change. Then write the rival claims the literature says are plausible.

Before data collection.

2

Define the evidence

For each hypothesis, name the observation that would support it — and the one that would count against it.

Before you see results.

3

Weigh and conclude

Compare how much evidence there is for each hypothesis, and say plainly what you can and cannot claim.

Judgement, made transparent.

CASE STUDY

Evaluation of the Embrace Multicultural Mental Health Project

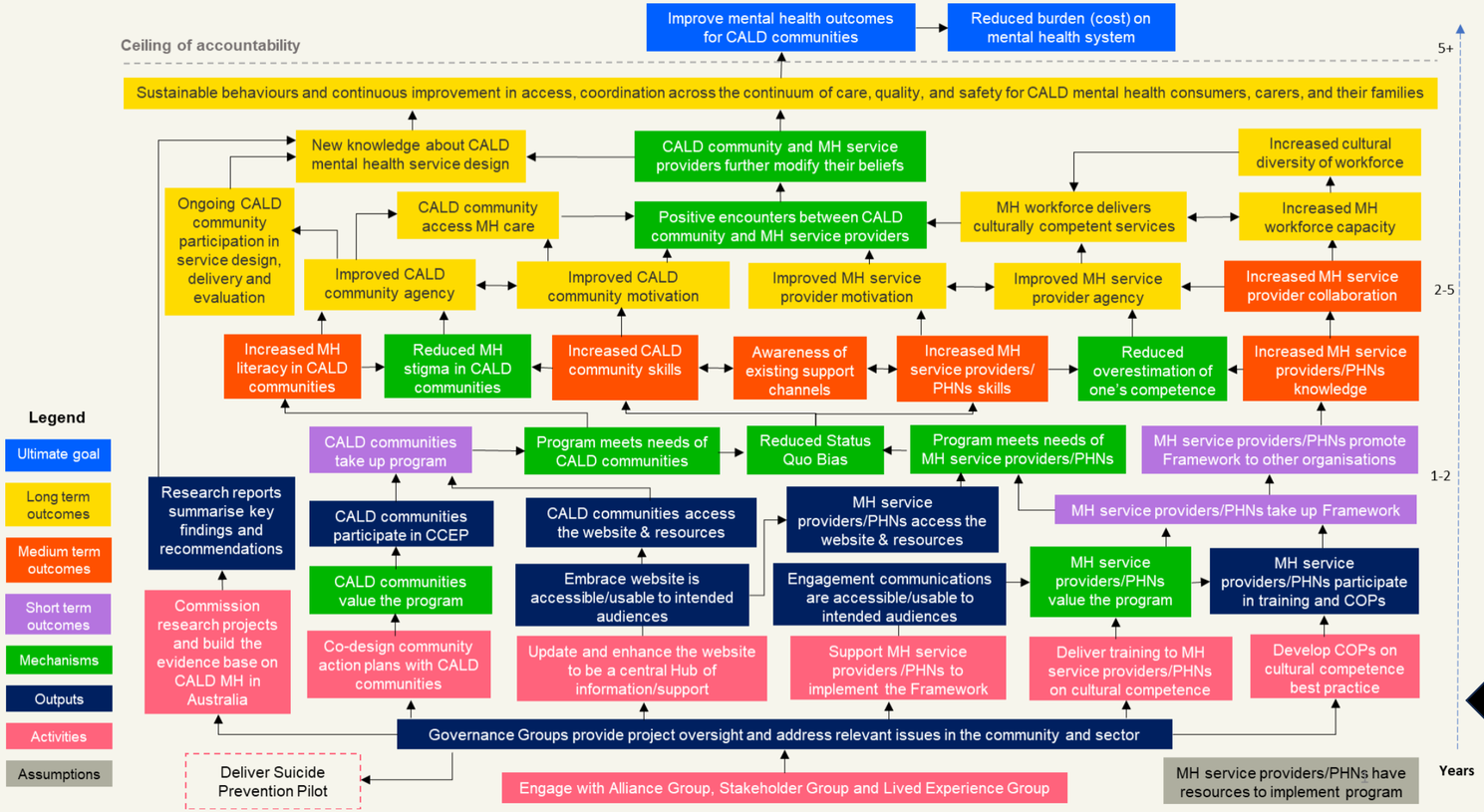
A national project to improve the quality and accessibility of mental health services for people from culturally and linguistically diverse backgrounds. Funded by the Department of Health and Aged Care, delivered by Mental Health Australia.

Verian evaluated it against nine key evaluation questions and reported in June 2024.



CASE STUDY

The Embrace Project Theory of Change



The ToC led to 9 KEQs about the project's appropriateness, implementation, outcomes, and cost-effectiveness.

KEY PROJECT COMPONENT FOR TODAY'S WORKSHOP

Framework for Mental Health in Multicultural Australia

Supports mainstream mental health services and practitioners, and PHNs who commission services, to assess their cultural responsiveness and enhance delivery of services for CALD communities.

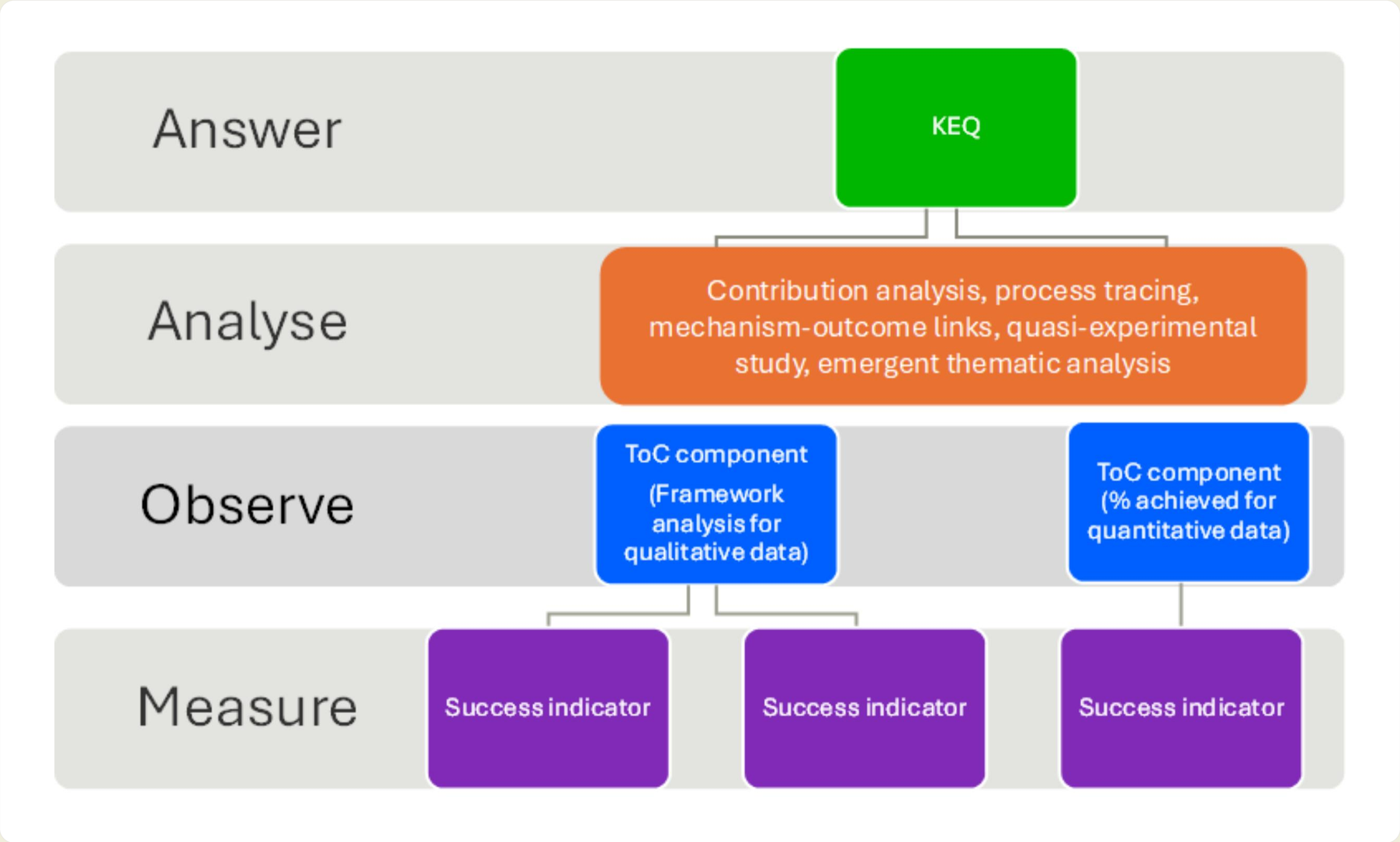
CASE STUDY

Data Sources

	Qualitative data				Quantitative data	
Data collection method	Interviews	Ethnography	Usability testing	Performance Reports	Survey	Admin data
Data analysis method	Emergent thematic analysis Framework analysis			Emergent thematic analysis Framework analysis	Process tracing · Contribution analysis · Framework analysis · Emergent thematic analysis · Quasi experimental	Count Emergent thematic analysis
Sample	Governance group members, PHNs, Service Providers, CCEP Orgs, MHA rep, DoHAC rep	Verian evaluators	MH service provider communications lead	MHA activity workplans, PHN engagement reports, CCEP progress reports, 2016 Review	Online survey of PHNs and mental health service providers · Pop-up HotJar poll on the usability of the Embrace website	Embrace budget, website and social media data

CASE STUDY

Data Matrix Approach



CASE STUDY

Sample size

210

PHN and service provider survey responses

With an embedded quasi-experiment

26

In-depth stakeholder interviews

Department, delivery, governance, PHNs

36

Project documents reviewed

Work plans, budgets, performance reports

15

Ethnography, usability and UX participants

Website, Framework and newsletter

1

Rapid evidence review

Source of the alternative hypotheses

Limitation: 210 usable responses from a sample of approximately 9,200 — a 2.3% response rate — with the usual risks of response and sample bias in self-reported, non-random data.

EXERCISE 1

From theory of change to testable hypotheses

5 min solo

5 min plenary

Sax Institute

EXERCISE 1 · YOUR TURN

Write one core hypothesis and two alternatives

ONE STRAND OF THE EMBRACE THEORY OF CHANGE

Embrace publishes the Framework

↓

The workforce registers and works through the modules

↓

They reflect on their own cultural competence and become more self-aware — mechanism

↓

Workforce cultural competence improves

The core hypothesis

Read it straight off the ToC pathway. How is cultural competence theorised to improve?

Two alternative hypotheses

What else could produce improved cultural competence in the MH workforce, with no help from Embrace at all?

EXERCISE 1 · WHAT THE EVALUATION DID

The hypotheses Verian tested

We only created process tracing hypotheses for priority KEQs about the impact of the Embrace Framework and CALD Community Engagement Project (CCEP).

KEQ	CORE HYPOTHESIS	ALTERNATIVE HYPOTHESIS
KEQ5	The CCEP reduced mental health stigma	Mental health stigma has reduced in CALD communities' home countries.
KEQ6	**Cultural competence is higher among PHNs and service providers that have used the Embrace Framework vs those that haven't** ◀ THE HYPOTHESIS WE EXAMINE TODAY	**Cultural competence is higher among PHNs and service providers that used other forms of professional learning.**
KEQ6	At least 70% of PHNs and service providers who have registered for the Embrace Framework agree with these statements: The Embrace Framework has helped me evaluate my cultural responsiveness. The Embrace Framework has helped me enhance my cultural responsiveness.	The workforce believes their practices became more culturally responsive because of something else.
KEQ6	The workforce believes their practices became more culturally responsive because of greater agency/ motivation/ capacity/ diversity.	The workforce believes their practices became more culturally responsive because of something else.
KEQ7	Improvement in CALD community access to mental health services is higher among PHNs and service providers that have engaged with the Embrace Project vs those that haven't.	Something else is driving improvement in CALD community access to mental health services.
KEQ7	Improvement in CALD community access to mental health services is associated with increased availability of information and reduced costs and stigma.	Something else is driving improvement in CALD community access to mental health services.

EXERCISE 2

Designing evidence tests that separate the hypotheses

5 min solo

5 min plenary

EXERCISE 2 · YOUR TURN

What would each hypothesis predict in the data?

CORE

Cultural competence is higher among PHNs and service providers that have used the Embrace Framework vs those that haven’t

ALTERNATIVE

Cultural competence is higher among PHNs and service providers that used other forms of professional learning.

If the core hypothesis is true…

Name one observation you would expect to find. Which data source gives it to you?

If the alternative is true…

Name one observation you would expect instead.

The mind-changer

What observations would provide disconfirming evidence for these hypotheses?

EXERCISE 2 · WHAT THE EVALUATION DID

Each hypothesis had its own evidence test

CORE HYPOTHESIS

Cultural competence is higher among PHNs and service providers that have used the Embrace Framework vs those that haven’t

EVIDENCE TEST

Quasi-experiment inside the survey: registrants versus non-registrants, plus a dose–response test on modules completed

WHAT WOULD COUNT

A significant competence gap between the groups, rising with dose

ALTERNATIVE HYPOTHESIS

Cultural competence is higher among PHNs and service providers that used other forms of professional learning.

EVIDENCE TEST

One multiple-choice question — “which organisation best supports your learning needs?” — with options randomised to avoid question order effects

WHAT WOULD COUNT

Higher competence among those naming a non-Embrace source

None of this required extra fieldwork. It required deciding, in advance, which question would separate the explanations.

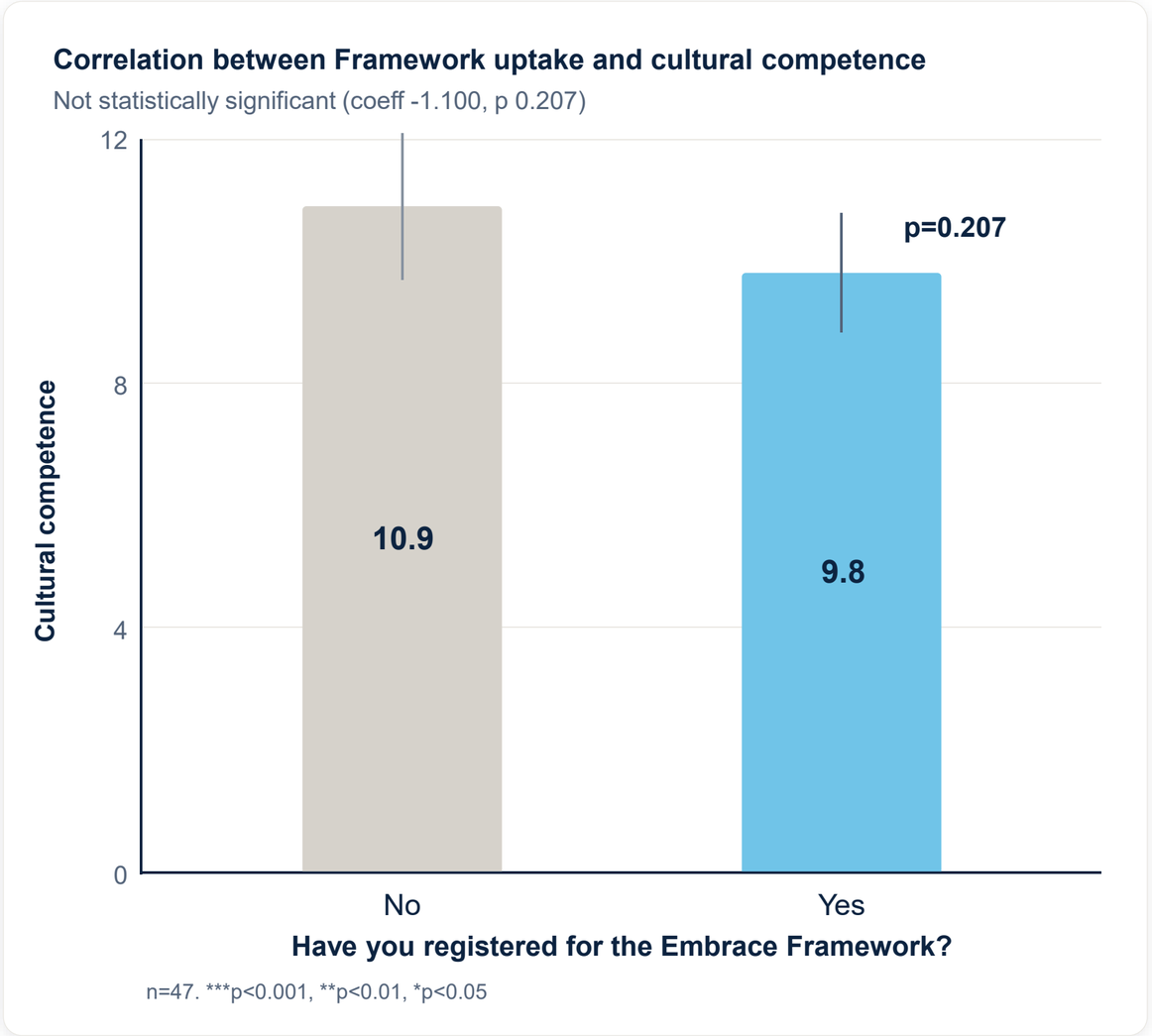
CASE STUDY

Core hypothesis: Cultural competence is higher among PHNs and service providers that have used the Embrace Framework vs those that haven’t

Cultural competence:
We created an index of three survey items about how often respondents say they do the following when interacting with people from a CALD background at work (including colleagues, delivery partners, or clients):

- I act to remove obstacles for people of different cultures when they identify such obstacles to me.
- I welcome feedback from people about how I can better relate to others from different cultural backgrounds.
- I find ways to adapt my work practices to people’s cultural preferences.

Our index comprises a sub-set of items from a validated cultural competence instrument reported in: Doorenbos, A. Z., Schim, S. M., Benkert, R., & Borse, N. N. (2005). Psychometric evaluation of the cultural competence assessment instrument among healthcare providers. *Nursing Research* , 54(5), 324-331.



CASE STUDY

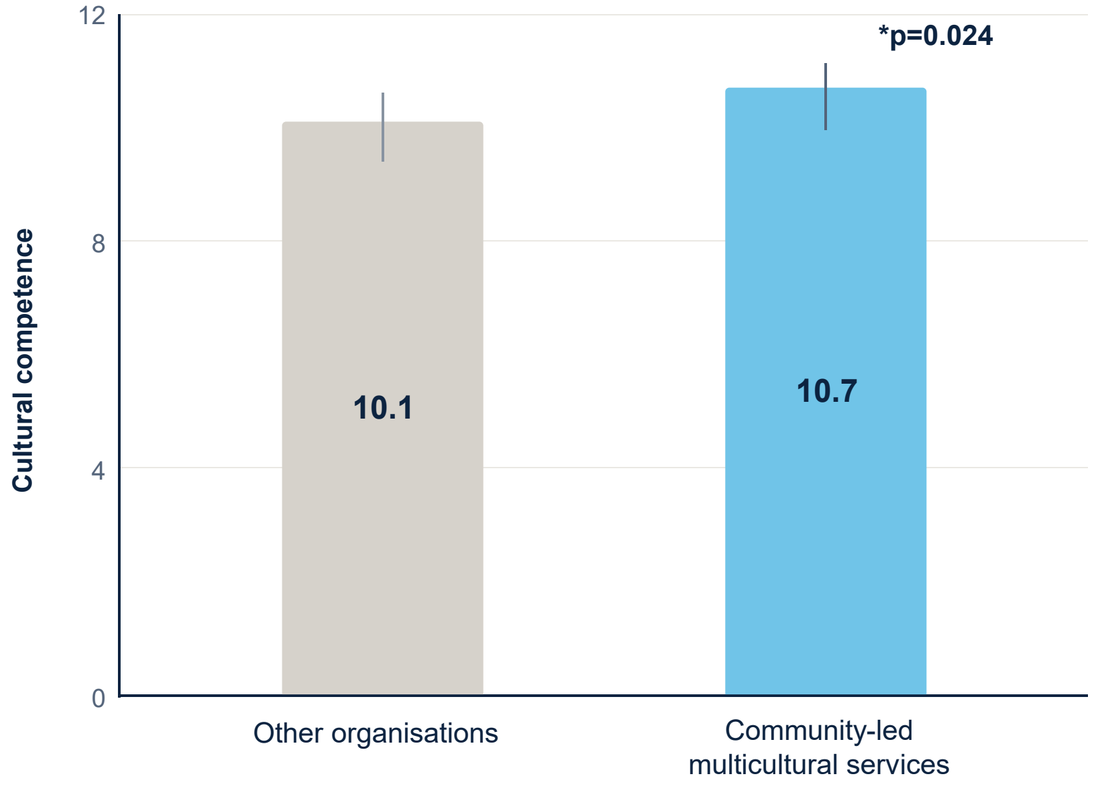
Alternative hypothesis: other sources of professional learning

Survey respondents were asked which organisation best supports their learning needs about culturally inclusive mental health service delivery. The following response options were randomly shown to avoid question order effects:

- Commonwealth Government e.g., Department of Health and Aged Care
- State Government agency
- **Community-led multicultural services**
- The Forum of Australian Services for Survivors of Torture and Trauma (FASSTT)
- Federation of Ethnic Communities' Council of Australia (FECCA)
- Embrace Multicultural Mental Health
- Other (please specify)

Those naming community-led multicultural services had higher cultural competence

Statistically significant (coeff 0.616, p 0.024)



Which organisation best supports your learning needs about culturally inclusive mental health service delivery?

n=149. ***p<0.001, **p<0.01, *p<0.05

CASE STUDY

A result that needed explaining

WHAT THE SURVEY SHOWED

No detectable effect of the Framework on workforce cultural competence.

Registering for the Framework was not associated with cultural competence (coeff -1.100, p 0.207).

WHAT THE INTERVIEWS SHOWED

Practitioners described the Framework changing how they worked.

Most PHN interviewees (n=5) reported they had improved their skills and knowledge from using the Framework.

Many evaluations encounter this. The usual resolution is to favour whichever data source the evaluator already trusts.

CASE STUDY

Three explanations fit the same pattern

EXPLANATION A

The program is not working

The Framework does not change practice. The positive interviews are social desirability bias or selection bias.

Implication: redesign or defund.

EXPLANATION B

The effects take time to materialise

Changing systems and processes is slow. Professional learning effects may need years, or a larger sample, to detect.

Implication: measure longer.

EXPLANATION C

Competing programs produce the same outcomes

The workforce is learning the same material elsewhere, so Embrace users and non-users look alike.

Implication: partner, or find the unmet gap.

Process tracing is the framework that lets you distinguish between them — and say which one the evidence actually supports.

EXERCISE 3

Weighing mixed evidence and reaching a defensible claim

7 min solo

5 min plenary

Sax Institute

EXERCISE 3 · YOUR TURN

Assign each to A, B, C or none.

B — No significant difference in cultural competence between Framework registrants and non-registrants (coeff -1.100, p 0.207).

C — Those naming a community-led multicultural service as best supporting their learning needs had higher cultural competence (coeff 0.616, p 0.024).

C — Just over half of all respondents (54%, n=74) said a community-led multicultural service best supports their learning needs, followed by Embrace (16%, n=22).

A — Framework registration was associated with a 22% improvement in self-awareness of own cultural competence (coeff -1.633, p 0.035).

B — The more recently respondents had accessed the Framework, the higher their cultural competence (coeff 0.398, p 0.051).

A — Most PHN interviewees (n=5) said they had improved their skills and knowledge from using the Framework.

Evidence can be confirming, neutral, or disconfirming of a hypothesis.

A The program is not working

B The effects take time to materialise

C Competing programs produce the same outcomes

EXERCISE 3 · WHAT THE EVALUATION CONCLUDED

How the three explanations fared

EXPLANATION A

The program is not working

Not supported

Findings 4 and 6 are disconfirming. The mechanism of change was activated in Framework registrants, and interviewees reported improved skills.

HOOP TEST · FAILED

The necessary confirmatory evidence was not observed, so A can be eliminated.

EXPLANATION B

The effects take time

Cannot be resolved

Findings 1 and 5 are neutral. Cultural competence was not associated with Framework registration, but recent users had higher cultural competence.

STRAW-IN-THE-WIND · PASSED

The evidence is weak either way — not confirming or disconfirming.

EXPLANATION C

Competing programs produce the same outcomes

Supported

Findings 2 and 3 are confirming. Community-led multicultural services are both the most-used and the most effective learning source.

SMOKING GUN · PASSED

Strong confirmatory evidence for C, but B is not strongly disconfirmed.

THE DEFENSIBLE CLAIM

Embrace contributed to the workforce accurately estimating its own cultural competence. It cannot be credited with a unique contribution to cultural competence itself, in a market where community-led services do the same work — and the evaluation design cannot rule out a longer time lag.

REFLECTION

Process tracing does not remove judgement — it makes it transparent

CASE STUDY

How the evidence informed recommendations

BECAUSE THE EVIDENCE POINTED HERE	THE RECOMMENDATION WAS
Community-led multicultural services are the dominant and most effective learning source	Explore delivering the Embrace Project in partnership with community-led multicultural services, and identify market gaps to develop new offerings
The Framework narrowed the difference between perceived and actual cultural competence (reduced Dunning-Kruger Effect)	Measure the difference between actual and perceived cultural competence as part of the Framework, and introduce more ways to increase the workforce’s self-awareness
Some aspects of the theory of change could not be adequately tested with the available data	Embed ongoing monitoring and evaluation in the project design for more rigorous evaluation in the future and to support continuous improvement

DISCUSSION

Taking this back to your own evaluation

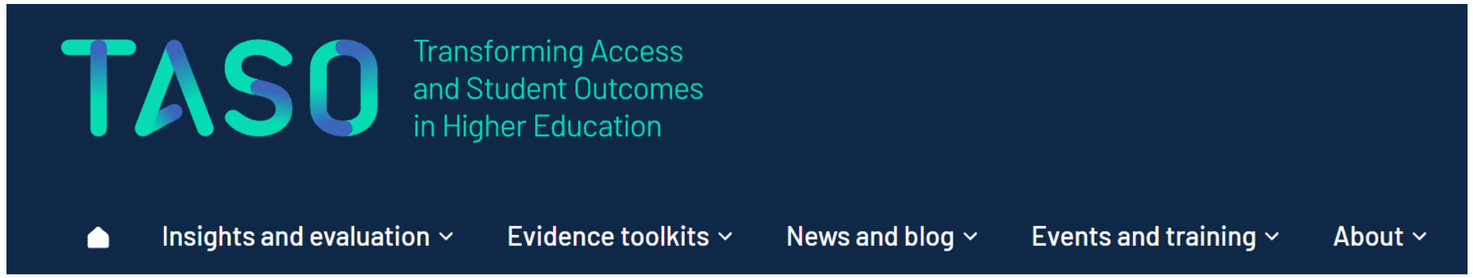
1. In the evaluation you are running now, what is the rival explanation you have not written down?

2. Could your current design see it if it were true — or would it look exactly like success?

3. What one question, added now, would separate the two explanations?

Process tracing is most affordable at design stage. Adding one well-chosen question before fieldwork is cheaper than defending an ambiguous finding after it.

Learning resource for further reading



Guidance and resources

Process tracing

Process tracing: the analysis of evidence on processes, sequences, and conjunctures of events within a case for the purposes of either developing or testing hypotheses about causal mechanisms that might causally explain the case.

THANK YOU

Questions, and where to read the case in full

PRESENTER

Dr Kizzy Gandy

Head, Evidence and Impact
Sax Institute
kizzy.gandy@saxinstitute.org.au

THE CASE STUDY

Verian, *Evaluation of the Embrace Multicultural Mental Health Project* , prepared for the Department of Health and Aged Care, 28 June 2024.

https://www.health.gov.au/resources/publications/evaluation-of-the-embrace-multicultural-mental-health-project-final-report?language=en