MEKSI

THE RESEARCH BEHIND THE BUILD

Making clinical reasoning visible

What the published literature says has to happen for clinical reasoning to be taught and assessed, and how MEKSI has tried to build each of those steps.

Dr Jawahar Thomas MBBS FRACGP · Founder, MEKSI

September 2026 · meksi.com

THE PATHWAY

1. INVISIBLE REASONING
2. VISIBLE PERFORMANCE
3. IDENTIFY GAPS
4. TARGETED FEEDBACK
5. REPEATED PRACTICE
6. MEASURABLE CHANGE

Six steps. Each one has a literature behind it.

BEFORE ANYTHING ELSE

What this deck is, and what it isn't

WHAT IT IS

- A map of the published work that shaped how MEKSI was designed.
- A statement of the mechanism we think matters: reasoning you can see is reasoning you can teach.
- An invitation to test whether that mechanism holds.

WHAT IT ISN'T

- Evidence that MEKSI works. There isn't any yet.
- A claim that these researchers endorse MEKSI. As far as we know, none of them has heard of it.
- A product demonstration. That is what the seventy-second film is for.

Nobody outside MEKSI has evaluated any of this. The last slide is an offer to pay for somebody to.

THE PROBLEM

Diagnostic error is a leading cause of preventable harm

~10% of patient deaths have a diagnostic error contributing, across decades of post-mortem research.

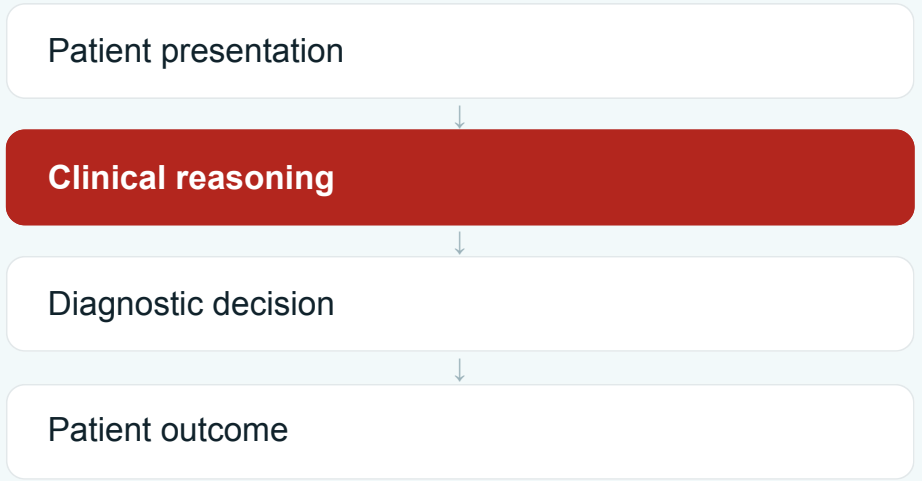
5% of US adults who seek outpatient care experience one in any given year.

Most people will experience at least one diagnostic error in their lifetime.

NASEM 2015 | Scott 2009 (BMJ) | Scott & Crock 2020 (MJA)

National Academies of Sciences, Engineering and Medicine, Improving Diagnosis in Health Care (2015).

WHERE IT ENTERS



Cognitive factors are primary or contributory causes of more than 75% of diagnostic errors. Knowledge deficits cause fewer than 5%.

THE TARGET

Clinical reasoning is four things, and each fails differently

1

GATHER

information

Fails as: the question that was never asked.

→

2

INTERPRET

evidence

Fails as: the finding noticed, then set aside.

→

3

FORMULATE

diagnoses

Fails as: the differential that closed too early.

→

4

DECIDE

management

Fails as: right diagnosis, wrong next step.

A single final mark cannot tell you which of the four went wrong. That is the assessment problem, not a scoring preference.

Scott 2009 · Scott & Crock 2020
Parsons et al. 2024

WHY IT IS HARD TO TEACH

Most of clinical reasoning never surfaces

“Complex, tacit and effectively invisible for students.”

Delany & Golding, BMC Medical Education, 2014,
on why clinical reasoning resists teaching.

Their words, about a study of twenty-one experienced allied health educators in three Australian teaching hospitals. What a supervisor observes is the output. The process that produced it stays inside the learner's head, where it can be neither corrected nor credited.

You can mark the answer.
You cannot mark the thinking behind it.

VISIBLE

Diagnosis · Decision · Outcome

INVISIBLE

- Information gathering
- Hypothesis generation
- Interpretation of findings
- Differential diagnosis
- Management reasoning

THE ARGUMENT IN ONE SLIDE

Six steps, and each one has a literature behind it

1 **INVISIBLE REASONING**
Delany & Golding 2014

→

2 **VISIBLE PERFORMANCE**
Schaye et al. 2025

→

3 **IDENTIFY GAPS**
Parsons et al. 2024

→

4 **TARGETED FEEDBACK**
Guerrasio & Aagaard 2014

→

5 **REPEATED PRACTICE**
Scott & Crock 2020
Waechter et al. 2022

→

6 **MEASURABLE CHANGE**
Staal et al. 2023
Waechter et al. 2026

The steps are sequential, not a menu. If reasoning is never made visible, steps three to six are simply not available to an educator, however good their intentions. The next six slides take one step each: what the research found, and what MEKSI does about it.

1 → 6

Break the chain anywhere and the rest cannot happen.

STEP 1 OF 6 · INVISIBLE REASONING

You cannot teach what you cannot see

WHAT THE RESEARCH FOUND

Delany and Golding worked with twenty-one experienced allied health clinical educators across three Australian teaching hospitals, trying to make expert clinical reasoning teachable. Their starting problem, in the paper's own words, is that clinical reasoning is “complex, tacit, and effectively invisible for students”.

Their remedy was to have each educator surface their own reasoning as an explicit heuristic and trial it with students. Note the direction: that paper makes the teacher's thinking visible. MEKSI runs the same principle the other way, on the learner's. The obstacle it starts from is the same one.

Delany & Golding 2014, BMC Med Educ 14:20

Nothing further along in this deck makes sense until this first problem is taken seriously.

WHAT MEKSI DOES HERE

Nothing. This step is the problem statement, not a feature.

It is on the slide because every design decision that follows is an answer to it, and a reader is entitled to see the question before the answers.

It is also the only slide in this deck with no product on it.

INVISIBLE REASONING | VISIBLE PERFORMANCE | IDENTIFY GAPS | TARGETED FEEDBACK | REPEATED PRACTICE | MEASURABLE CHANGE

STEP 2 OF 6 · VISIBLE PERFORMANCE

Reasoning becomes assessable the moment it is performed

WHAT THE RESEARCH FOUND

Schaye and colleagues took the principle to written clinical work. Across two medical schools they built and validated named-entity and large-language-model assessments of clinical reasoning documentation, scoring roughly 1,400 resident admission notes against a previously validated rubric.

Their own conclusion is narrower than ours: such tools can enhance feedback on clinical reasoning, and the approach generalised between the two sites. The leap to scale is our inference, not theirs. They also report, plainly, that some rubric dimensions performed well and others were insufficient for implementation.

Schaye et al. 2025, JMIR 27:e67967

WHAT MEKSI DOES

- A learner interviews a simulated patient in free text. There is no menu of options to pick from, so nothing is prompted.
- Every question asked, every question not asked, the order they came in, and what the learner did with each answer is captured as it happens.
- Nothing is reconstructed afterwards from memory, by the student or by the supervisor.

The consultation is not a test wrapped around the reasoning. It is the reasoning, written down.

STEP 3 OF 6 · IDENTIFY GAPS

A single score hides the thing you actually need to fix

WHAT THE RESEARCH FOUND

Parsons and colleagues took 114 residents and fellows referred over five years for performance concerns of any kind. In thirty-eight of them, a third, the deficiency was clinical reasoning.

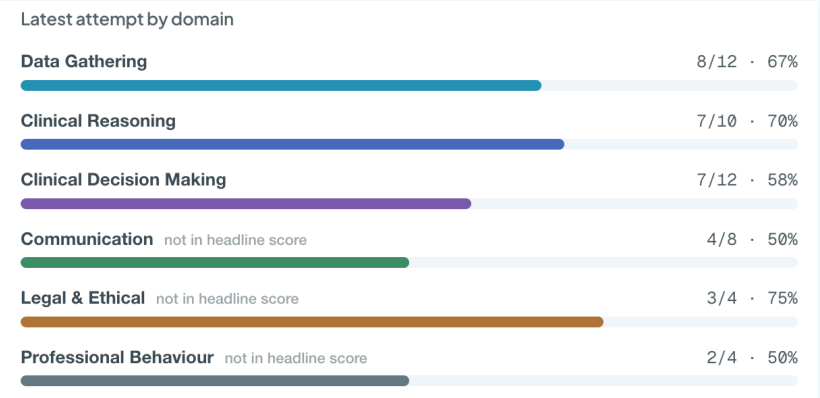
Targeted assessment then located what each one's deficit actually was: hypothesis generation in 42 per cent, problem representation in 18 per cent, data gathering in 16 per cent, management in 16 per cent, hypothesis refinement in 8 per cent. Five different problems wearing one label. No composite mark can separate them, and no remediation plan written against the label can fix them.

Parsons et al. 2024, J Grad Med Educ 16(4)

“Weak clinical reasoning” is not a finding. “Closed the differential after two questions” is.

WHAT MEKSI DOES

Per domain, per attempt. Never one case mark.



A learner's latest attempt, by rubric domain.

INVISIBLE REASONING | VISIBLE PERFORMANCE | **IDENTIFY GAPS** | TARGETED FEEDBACK | REPEATED PRACTICE | MEASURABLE CHANGE

STEP 4 OF 6 · TARGETED FEEDBACK

Feedback changes behaviour only when it points at something specific

WHAT THE RESEARCH FOUND

Guerrasio and Aagaard took 151 referrals to a remediation programme over six years. Fifty-three had clinical reasoning deficits. Each was given a standardised ten-step plan built against their specific deficit, and fifty-one of the fifty-three passed independent reassessment.

The finding that matters here is the price. Remediation took a median of eighteen hours of faculty time per learner, and up to a hundred. Targeted feedback works. Delivering it by hand does not scale.

Guerrasio & Aagaard 2014, J Gen Intern Med 29(12)

WHAT MEKSI DOES

- Feedback is written against the omission, not against the score.
- It names the question that was not asked, what the answer would have changed, and where in the consultation it should have come.
- The learner can see the point in their own transcript where the path diverged.

Feedback inherits the resolution of whatever assessment sits underneath it.

INVISIBLE REASONING | VISIBLE PERFORMANCE | IDENTIFY GAPS | **TARGETED FEEDBACK** | REPEATED PRACTICE | MEASURABLE CHANGE

STEP 5 OF 6 · REPEATED PRACTICE

Reasoning improves with deliberate repetition, not with exposure

WHAT THE RESEARCH FOUND

Scott and Crock's prescription is specific: deliberate practice in which the clinician works diagnostic cases while verbalising their reasoning, then compares that reasoning against an expert's. They add the reason it has to be external. Clinicians' self-assessment of their own diagnostic accuracy is unreliable, so the comparison cannot come from the learner.

Waechter and colleagues built an instrument that does it: a data-rich virtual-patient tool that scores the reasoning itself and returns individualised feedback. It is the platform behind both studies on the next slide.

Scott & Crock 2020, MJA 213(7)

Waechter et al. 2022, Acad Med 97(10)

WHAT MEKSI DOES

- Cases are grouped so that one reasoning step can be practised across several different presentations.
- A learner who closes the differential too early gets four more chances to do it differently, this week rather than next rotation.
- Repetition is scheduled by the educator, rather than left to the roster.

The cases a student needs are rarely the cases that walk in.

INVISIBLE REASONING | VISIBLE PERFORMANCE | IDENTIFY GAPS | TARGETED FEEDBACK | **REPEATED PRACTICE** | MEASURABLE CHANGE

STEP 6 OF 6 · MEASURABLE CHANGE

Repeat the performance and the change becomes measurable

WHAT THE RESEARCH FOUND

Two studies, on the same virtual-patient platform. Staal and colleagues followed 121 pre-clerkship students through ten to eleven simulated cases with expert scorecards and individualised feedback. Diagnostic justification scores rose 40 per cent, from a poor base: students scored 30 to 48 per cent lower on justifying a diagnosis than on reaching one.

Waechter and colleagues then ran it across five North American medical schools — 1,066 students, 12,400 cases, 1.55 million datapoints — and the gains held at scale, with no plateau after twenty cases.

Staal et al. 2023, BMC Med Educ 23:684

Waechter et al. 2026, BMC Med Educ 26:414

The thing that improved is the thing that had been made visible.

WAECHTER 2026 · WITH REPEATED PRACTICE

Diagnostic justification	+72%
Test ordering	+40%
Misdiagnoses	3x fewer
Cognitive errors	~half

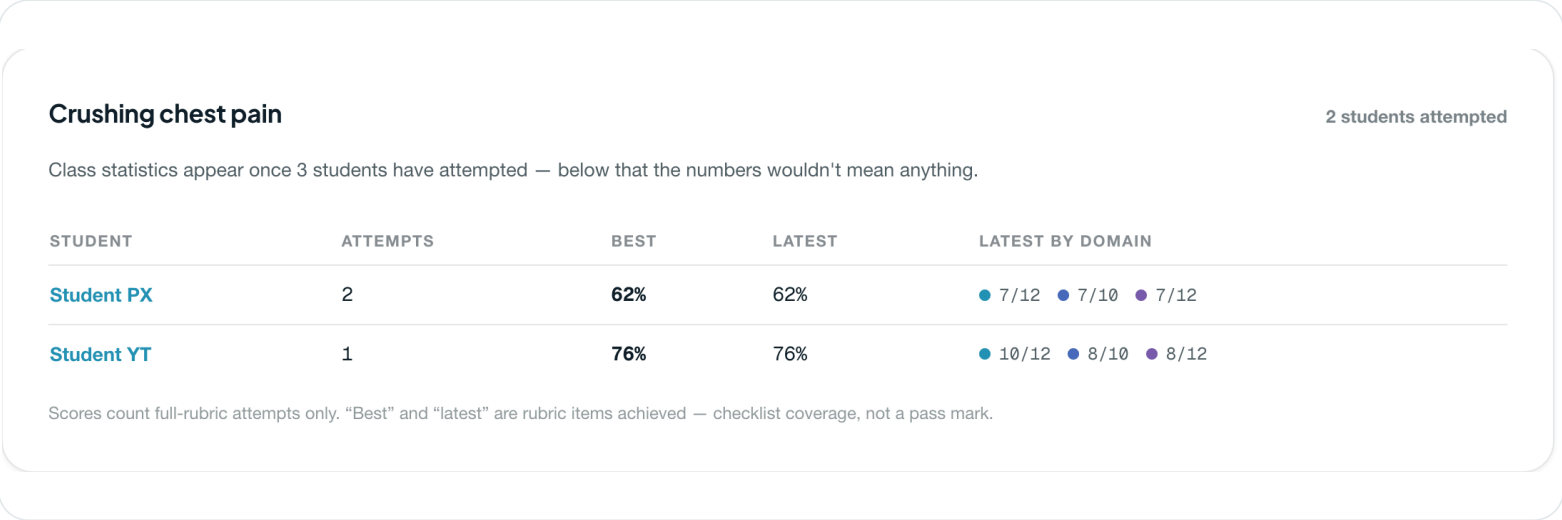
1,066 students · 12,400 cases · five schools

Staal 2023: justification +40% after 10–11 cases

INVISIBLE REASONING · VISIBLE PERFORMANCE · IDENTIFY GAPS · TARGETED FEEDBACK · REPEATED PRACTICE · **MEASURABLE CHANGE**

WHAT IT LOOKS LIKE IN USE

The educator sees the class, not the individual



Crushing chest pain — 2 students attempted

Class statistics appear once 3 students have attempted — below that the numbers wouldn't mean anything.

STUDENT	ATTEMPTS	BEST	LATEST	LATEST BY DOMAIN
Student PX	2	**62%**	62%	7/12 · 7/10 · 7/12
Student YT	1	**76%**	76%	10/12 · 8/10 · 8/12

Scores count full-rubric attempts only. “Best” and “latest” are rubric items achieved — checklist coverage, not a pass mark.

The cohort view in the shipping product, against a seeded demonstration class.

The privacy design is not a compliance afterthought. It is what makes the data usable in a teaching setting at all.

Pseudonymous by design

Students appear under a stable neutral label that never maps to a name or an email in this view.

Statistics are gated

Class figures appear only once three students have attempted. Below that the numbers would not mean anything.

Per domain, not per mark

Best, latest, and a breakdown across data gathering, reasoning and decision making.

Not a grade book

For graded assessment the institution uses its own LMS. This view exists so teaching can be adjusted.

REFERENCES

Everything cited, in full

1. **National Academies of Sciences, Engineering, and Medicine.** Improving Diagnosis in Health Care. Washington DC: The National Academies Press; 2015.

2. **Scott IA.** Errors in clinical reasoning: causes and remedial strategies. BMJ. 2009;338:b1860. doi:10.1136/bmj.b1860

3. **Scott IA, Crock C.** Diagnostic error: incidence, impacts, causes and preventive strategies. Med J Aust. 2020;213(7):302–305.e2. doi:10.5694/mja2.50771

4. **Delany C, Golding C.** Teaching clinical reasoning by making thinking visible: an action research project with allied health clinical educators. BMC Med Educ. 2014;14:20. doi:10.1186/1472-6920-14-20

5. **Schaye V, DiTullio D, Guzman BV, et al.** Large language model-based assessment of clinical reasoning documentation in the electronic health record across two institutions: development and validation study. J Med Internet Res. 2025;27:e67967. doi:10.2196/67967

6. **Parsons AS, Dreicer JJ, Martindale JR, Young G, Warburton KM.** A targeted clinical reasoning remediation program for residents and fellows in need. J Grad Med Educ. 2024;16(4):469–474. doi:10.4300/JGME-D-23-00822.1

7. **Guerrasio J, Aagaard EM.** Methods and outcomes for the remediation of clinical reasoning. J Gen Intern Med. 2014;29(12):1607–1614. doi:10.1007/s11606-014-2955-1

8. **Waechter J, Allen J, Lee CH, Zwaan L.** Development and pilot testing of a data-rich clinical reasoning training and assessment tool. Acad Med. 2022;97(10):1484–1488. doi:10.1097/ACM.0000000000004758

9. **Staal J, Waechter J, Allen J, Lee CH, Zwaan L.** Deliberate practice of diagnostic clinical reasoning reveals low performance and improvement of diagnostic justification in pre-clerkship students. BMC Med Educ. 2023;23:684. doi:10.1186/s12909-023-04541-5

10. **Waechter J, Kusnoor A, Eickman K, et al.** Reducing misdiagnoses and cognitive errors using virtual patients and automated feedback in a clinical reasoning curriculum. BMC Med Educ. 2026;26:414. doi:10.1186/s12909-026-08647-4

Declared conflict, reference 10: Jason Waechter owns teachingmedicine.com, the platform that generated the data in references 8, 9 and 10.

LIMITATIONS

What none of this establishes

01

No independent evaluation

Nobody outside MEKSI has evaluated the platform. Everything in this deck is design rationale.

02

An internal scoring framework

The rubric is expert-derived and internally reviewed. It has never been externally validated.

03

Simulated is not real

Performance with a simulated patient may not predict performance with a real one.

04

Improvement on our own cases

Until it is checked against an independent outcome, that is all a rising MEKSI score means.

Each of those four is a study somebody could run. The next slide is an offer to fund them.

THE OFFER

A funded pilot, designed and run by your team

1

Content validation

Clinicians review the cases, the expected reasoning and the scoring framework.

→

2

Assessor validation

Does MEKSI identify reasoning behaviours and omissions a clinician agrees are meaningful?

→

3

Student pilot

A defined cohort completes a defined set of cases.

→

4

Repeated practice

Does visible feedback plus repetition produce measurable improvement?

→

5

External comparison

Compare MEKSI performance against independent educator or real-world assessment.

MEKSI PROVIDES

Platform and licences · the existing case library · new cases written to your curriculum · the funding for the study

YOUR TEAM OWNS

The research question and method · the analysis · the conclusions, including negative ones · publication, with no right of review or veto from us

MEKSI

What would it take to test this properly?

That is the whole question, and it is genuinely the one we would like to be asked. Twenty minutes is enough to work out whether a pilot is worth designing.

Dr Jawahar Thomas MBBS FRACGP

Founder, MEKSI · General practitioner, South Australia

info@meksi.com

meksi.com · the 70-second film is in the covering email