Choosing What to Keep Human

Assessing Competence When AI Shares the Work

Carl Preiksaitis, MD MEd · Stanford Emergency Medicine
IU School of Medicine · Annual Program Directors Meeting

STANFORD MEDICAL CENTER

HISTORY AND PHYSICAL / ADMISSION NOTE

Page ____ of ____

DATE 5/14/88 TIME 1845 SERVICE Med. ROOM 512-B

PATIENT _____ AGE 49 SEX M

MR# _____ DOB _____

ADMITTING PHYSICIAN R.L. Martin, M.D.

CONSULTS _____

CHIEF COMPLAINT Chest pain.

HISTORY OF PRESENT ILLNESS

49 y/o WM. c/o substernal chest pressure since this AM... Began while mowing lawn ~0900. Pressure has been intermittent, lasting 10-15 min., associated with SOB and diaphoresis. No radiation. No nausea or vomiting. Sx. relieved with rest.

No prior episodes. No recent travel. Denies cough, fever, or sputum.

? exertional component vs. pleuritic

PAST MEDICAL HISTORY HTN x 5 yrs. Hyperlipidemia.

PAST SURGICAL HISTORY Appendectomy (age 12).

MEDICATIONS HCTZ 25 mg qd. Atenolol 50 mg qd.

ALLERGIES NKDA

FAMILY HISTORY Father d. MI age 62. Mother A&W.

SOCIAL HISTORY 1 ppd x 25 yrs. Occas. ETOH. Works as mech. Married, 2 children.

REVIEW OF SYSTEMS as above, otherwise negative.

PHYSICAL EXAM

BP (148/88) P 96 R 20 T 98.6

GEN: WDWN male, anxious, diaphoretic.
HEENT: WNL
NECK: Supple, no JVD, no bruits.
CHEST: Clear bilat.
CV: RRR, no murmurs, rubs, or gallops.
ABD: Soft, NT, no HSM.
EXT: No c/c/e. Distal pulses 2+.
NEURO: Non-Focal.

Pain now 2/10 after 2 SL NTG.
Improved after NTG.
Any change w/ inspiration?
? GI source

ASSESSMENT

1. Chest pain - (r/o ischemia) (↑ risk factors)
2. ~~GERD / esophageal spasm~~ [consider, but less likely]
3. ~~Musculoskeletal pain~~ [no reproducible tenderness]
4. PE (unlikely) [no risk factors]

Most likely: Unstable Angina

Ddx:
- (ACS / UA)
- ~~GERD / esoph spasm~~
- ~~Musculoskeletal~~
- PE (unlikely)
- Aortic dissection (unlikely) [no tearing pain, pulses equal]

PLAN

1. Admit to telemetry.
2. O2 2 L/min.
3. ASA 325 mg chewed.
4. NTG paste 1" q6h.
5. EKG now and in AM.
6. Cardiac enzymes q6h x3.
7. NPO after MN.
8. Stress test vs. cath pending enzymes.

[If enzymes + or recurrent pain → cath]

PHYSICIAN SIGNATURE R. L. Martin, M.D. DATE 5/14/88 TIME 1905

* follow enzymes / ECGs closely

Thorough. Well organized. Efficient. Ready.

Who produced the evidence we are using to judge the resident?

AI can improve the finished work while making the resident's reasoning harder to see.

THE WORK YOU USE TO JUDGE A LEARNER

A preoperative plan

An image interpretation

A psychiatric formulation

Where does the learner's judgment become visible?

Elderly patient, blunt trauma, image the neck.

Maybe that was the process. I could not tell.

WRITING THE NOTE

Gather

Select

Commit

Sequence

Revise

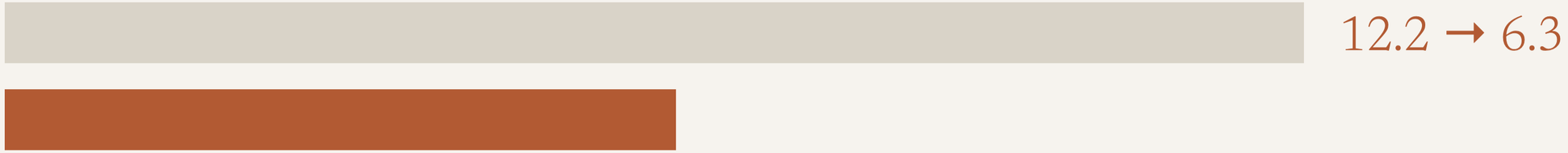
70

seconds saved per encounter

Ambient scribe use, on-shift documentation time, audit logs. Preiksaitis et al. JMIR AI. 2026;5:e92193.

TIME IN NOTES

REPORTED MENTAL DEMAND



Ambulatory evaluation, NASA-TLX. Stults et al. JAMA Netw Open. 2025;8(5):e258614.

THE NOTE, PRODUCED

Gather — Select — Commit — Sequence — Revise

Optional, and therefore invisible.

THE EDUCATIONAL WORK WITHIN THE TASK

Doing it can help a learner develop judgment.

Observing it can give faculty evidence of that judgment.

When the work changes, what does the evidence still support?

BEDSIDE	what faculty observe
FILE	what the committee receives
CLAIM	what the program certifies

~~Not detection.~~

Observation design.

Verify, then trust.

It becomes supervision when the learner is inside the loop.

Five supervision steps: Abdulnour, Gin & Boscardin. N Engl J Med. 2025;393(8):786–797. In the handout.

Walk me through the part you were least sure about.

HYPOTHETICAL CLINIC EXAMPLE

The summary says “fatigue.”

The patient says “weak.”

“What would you need to clarify?”

HYPOTHETICAL EXAMPLE · WHAT FACULTY OBSERVE

An open-ended question about uncertainty

Questions that pursue the discrepancy

An assessment revised with new information

The attending and resident agree on the plan together.

HYPOTHETICAL EXAMPLE · THE EVALUATION

“After an open-ended question about uncertainty, the resident recognized that she had not clarified what the patient meant by ‘weak.’ I observed her pursue that distinction through further history and a focused examination, then reconsider her assessment. We agreed on the plan together.”

One observed encounter, with the support recorded.

THE CIRCUMSTANCES OF THE OBSERVATION

What was the learner asked to do?

What help was present?

Assisted and unassisted describe the work. They are not grades.

INTERNAL MEDICINE MILESTONES 2.1 · PATIENT CARE 1: HISTORY

LEVEL 1	Acquires a history using a template-driven approach
LEVEL 2	**Elicits and concisely reports a hypothesis-driven patient history for common patient presentations**
LEVEL 3	Acquires and concisely reports a hypothesis-driven history in a patient with a complex clinical problem

121

accredited programs — and 121 places to ask what the evidence actually shows

Indiana University School of Medicine. 2025–2026 Fact Sheet.

You observed the system.

You are being asked about the person.

The physician must retain enough independent judgment to recognize, interrogate, and recover from consequential tool failure.

This individual can be trusted with unsupervised responsibility inside a complex system — including when one part of it is wrong.

What responsibility should we entrust to the tool?

A convincing explanation is a starting point.

The evidence needs to include performance.

SAME VENDOR

Generating practice questions

Error is cheap and visible

CONSIDERABLE ROPE

Ranking applications

Error is consequential and hard to detect

VERY LITTLE

Preiksaitis et al. SCALES-AI. AEM Educ Train. 2026;10(3):e70195.

An evaluation applies to a particular use.

Who will review it when that use changes?

The physician's judgment is the safeguard residency still has to certify.

01 Does doing this task help learners develop judgment?

02 Does it give faculty an important opportunity to observe that judgment?

WHAT WE STILL NEED TO LEARN

Whether these frameworks improve training

How AI changes readiness for independent practice

Which clinical activities develop judgment, and how often faculty must observe it

THIS YEAR

01 Run the milestone audit

THIS YEAR

01 Run the milestone audit

02 Give your faculty the question

THIS YEAR

01 Run the milestone audit

02 Give your faculty the question

03 Write down what you already let AI do

I am not saying this resident is not competent.

I am not saying this resident is not competent.

I am saying the file does not let me answer the question we are asking.

I am not saying this resident is not competent.

I am saying the file does not let me answer the question we are asking.

I have repeatedly observed good performance when the resident works with the tool.

I am not saying this resident is not competent.

I am saying the file does not let me answer the question we are asking.

I have repeatedly observed good performance when the resident works with the tool.

I have not directly observed the judgment this milestone asks me to rate.

I am not saying this resident is not competent.

I am saying the file does not let me answer the question we are asking.

I have repeatedly observed good performance when the resident works with the tool.

I have not directly observed the judgment this milestone asks me to rate.

Before I sign, we need an opportunity to observe that judgment.

She may be right that the file is insufficient.

She is not yet right about the resident.

Carl Preiksaitis, MD MEd

cpreiksaitis@stanford.edu

Handout: carlmd.com/keephuman