ical Skills and Assessment

rriculum Alignment and Evaluation

undation / Millennium Conference • Preclinical M1/M2 • November 2026

ument maps Session 2 learning objectives to teaching, practice, and assessment. It also provides an evaluation plan across Ki –4, a session survey, and an annotated reading list.

culum Alignment Matrix

slide numbers refer to the Session 2 lecture deck. PBL, Active Learning (AL), and MCQ references refer to the Session 2 delive ck levels are detailed in the Evaluation Plan below.

e (abbreviated)	Bloom's level	Lecture (slides)	PBL	Active Learning	MCQ	Evaluation
gnize and categorize AI the clinical task each	Remember / Understand	6–12	Case 1 Part 2; Case 2 Exhibits	Step 1 orientation	1	L2: MCQ pre/post
entiate AI-augmentable n core competencies; de-skilling / never-skilling	Analyze	14–16	Case 1 Parts 1 & 4	Prompt D	2	L2: MCQ; L3: observation car
dential/compliant vs. public nstitutional guidance; e and data protection	Apply	24	Case 1 Part 2; Section D	Prompt B	3	L2: MCQ; L3: self-audit
in how CDS generates ns; input quality, training umptions, limitations; in a specific case	Understand / Apply	8–9, 13	Case 1 Part 3 (Q5; Tier A re-run)	—	4	L2: MCQ pre/post
endent reasoning first; grate, question, or override with justification	Apply / Evaluate	17–18, 20	Case 1 Parts 1–3	Prompt A	5	L2: MCQ; L3: OSCE station
fy bias, obscured ty, reduced reasoning, de- sks; describe mitigations	Analyze / Evaluate	9, 15, 25	Case 1 Part 3; equity moments	Prompt D	2, 4	L2: MCQ pre/post

e (abbreviated)	Bloom's level	Lecture (slides)	PBL	Active Learning	MCQ	Evaluation
ion and limitations of stening / AI scribes	Understand	10–11	Case 2 Exhibits 1–4; Section D	Glossary	7	L2: MCQ pre/post
w, correct, validate AI e responsibility for errors, s, tone	Apply / Evaluate	21–22	Case 2 Tasks 1–2	Prompt B	6	L2: corrected-note checklist; L observation card; L4: docume signal
ate documentation AI's on attention, empathy, cation; preserving patient-care	Evaluate	21, 23	Case 2 Task 2 & debrief	—	7	L2: MCQ; L4: patient-experien signal
ured self-reflection using vely and transparently	Create	26	Case 1 Part 4; Case 2 Task 3	Entire activity (create-level)	5 (in part)	L2: boundary-statement rubric self-audit
ate AI output within ecision-making, balancing lues, judgment, ethics	Evaluate	27	—	—	8	L2: MCQ pre/post

the matrix. Every objective is taught (lecture), practiced (PBL and/or Active Learning), and assessed at Level 2 or higher. The activity is deliberately the sole create-level assessment: the boundary statement is an authored artifact, scored with the rubric an tested by MCQ. Objective 4b is practiced primarily through MCQ 8's shared-decision-making case and revisited in Session d controversy for schools running multiple sessions — but this session's coverage stands alone.

dary Statement Rubric (Level 2 artifact)

ach student's AI boundary statement 0–2 per criterion; 8 points total. Collected at the end of the Active Learning activity (or sub d).

	0 — Absent	1 — Named	2 — Specified
behavior	Intentions and values only (“I will be careful”)	At least one commitment framed as an action	Commitments an observer could v behaviors, not dispositions
families	Addresses neither or only one of CDS / documentation	Addresses both, but with one undifferentiated rule and no rationale	Addresses both and reasons expli about whether the rules should dif

	0 — Absent	1 — Named	2 — Specified
ecific risk	No risk named, or term misapplied	Correct term (de-, mis-, never-skilling) applied to self	Correct term plus a specific skill, a guardrail habit, and the situation o greatest temptation
check	No follow-up mechanism	A vague check-in (“I'll reflect on it”)	An observable three-month sign o boundary erosion or maintenance

ted competence threshold: 6/8 with no zero. Below-threshold statements identify students for brief follow-up; item-level patte weaknesses in the session itself.

ation Plan — Kirkpatrick Levels 1–4

y tiers accompany the higher levels so all eleven consortium schools can implement at least the core plan.

— Reaction

ession survey (below), about five minutes. Core signal: perceived ability on each objective cluster, value of the restructured ha nat, pacing.

2 — Learning

pre/post. The eight MCQs, administered before the session (or asynchronously as pre-work) and after. Paired scores give an ctive learning measure per objective cluster.
ndary-statement rubric. The create-level artifact, scored 0–8 as above.
ected-note checklist (Case 2). Groups submit their corrected note from Case 2, Task 2. A three-point checklist — identifies th ed omission (orthopnea thread), removes all four fabrications, adds the return-to-room questions — verifies documentation-vali n applied form.

3 — Behavior

E station, 3–6 months post-session. A standardized-patient encounter in which the student receives an AI-drafted note for th just conducted, seeded with one omission and two fabrications, and has five minutes to finalize it. Checklist: performs an indep ssment before consulting the draft; detects the seeded fabrications; corrects rather than annotates; obtains missing information P rather than inventing it; the signed product accurately reflects the encounter. Feasibility tier: schools without OSCE capacity ote-review task into an existing standardized-patient or documentation exercise.

kship direct-observation card. A brief EPA-style card for preceptors, completed once per student per relevant rotation: comm
pendent differential before consulting decision support; verifies AI-generated content before signing; can state where an AI
mmendation came from and why they accepted or overrode it. Aggregated, this is the curriculum's primary behavioral endpoint
ent self-audit (low-cost tier). At clerkship midpoint, students log one real encounter involving an AI tool and write three sente
they did independently first, what they verified, what they decided. Reviewed in existing mentoring structures; doubles as the
tudinal revision prompt for the boundary statement.

4 — Results

umentation-safety signal. Where note-audit programs exist, tag AI-drafted-note errors caught before versus after signature. T
roxy for verification behavior at scale.
s-preservation signal. Where clerkship OSCEs include a no-AI differential-diagnosis exercise, trend cohort performance year
A stable or rising trend against rising AI availability is the never-skilling countersignal this session exists to produce.
ent-experience signal. Where affiliated clinics use ambient scribes with student involvement, add one item to existing patient-
rience surveys: “My clinician gave me their full attention, and my concerns were reflected accurately in my visit summary.” Tren
line, is the metric.

on caveat. Level 4 outcomes are influenced by far more than one preclinical session. These indicators are monitored as conse
text, not claimed as session effects. A separate consortium evaluation working group should own this layer.

on Evaluation Survey

ms (1–5, strongly disagree to strongly agree)

name the major families of AI tools I will encounter in clinical training and the task each supports.
explain which clinical skills I must develop independently, and describe the risks of de-skilling and never-skilling in my own wo
state when AI use requires a secure, sanctioned tool and when publicly available tools are acceptable.
ng generated my own differential and then appraised an AI's version, I can describe how I will integrate, question, or override A
mmendations.
ng audited an AI-drafted note against a verbatim transcript, I can describe the specific checks I will perform before signing an A
rated note.
nands-on structure of the PBL cases (using and appraising real AI output) was a valuable use of class time.
ecture content was clearly presented and appropriately paced.

nce self-assessment (retrospective pre/post, 1–5)

aising an AI-generated differential for a specific patient, including what to verify and against what source.
ng the omissions and fabrications in an AI-drafted clinical note.

ded

t was the single most valuable thing you learned in this session?
t is one thing about this session that could be improved before it is taught again?

ed Reading List

owing readings support the session. Distinguish empirical findings from conceptual frameworks, and check current local poli
guidance.

tza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318(6):517–518. — T
, clinician-facing piece that names the deskilling and over-reliance failure modes this session's framework rests on. Excellent p
ng.

E, Gallo R, Hom J, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Network
;7(10):e2440969. — Evidence that access to an LLM does not automatically improve clinician diagnostic reasoning; integration
Grounds the reason-first workflow.

ey AA, Gayre G, Hoberman B, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. N
lyst. 2024. — Representative implementation report on ambient scribes; grounds the balanced framing of Case 2.

ecke A, Nam A, Lake E, et al. Racial disparities in automated speech recognition. PNAS. 2020;117(14):7684–7689. — Docu
racy variation across speaker groups; the evidence behind Case 2's equity teaching moment.

on RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for su
Digital Medicine. 2020;3:17. — Accessible CDS overview for students who want the tool taxonomy in more depth.

sell RG, Lovett Novak L, Patel M, et al. Competencies for the use of artificial intelligence–based tools by health care professio
emic Medicine. 2023;98(3):348–356. — The competency framework this session's skills map onto.

rican Medical Association. Professionalism in Health Care Systems. Code of Medical Ethics, Opinion 11.2.1. https://code-me
s.ama-assn.org/ethics-opinions/professionalism-health-care-systems — Institutional conditions for professional judgment, care
responsible use of decision-support systems.

zyń K, Romańczyk M, Kitala D, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a mult
rvational study. Lancet Gastroenterol Hepatol. 2025;10:896–903. doi:10.1016/S2468-1253(25)00133-5. — Lower unassisted a
ction was observed after AI exposure; the observational design does not establish that AI caused the decline or that this genera
nostic reasoning.

J, Handley JL, Cobb NK, et al. Accuracy and safety of AI-enabled scribe technology: instrument validation study. J Med Inter
;27:e64993. doi:10.2196/64993. — Simulated encounter study identifying omissions, additions, and wrong outputs. For deploy
atient experience, see Memon S, Brand A, Taylor B, et al. Performance, acceptability, and impact of ambient listening scribe
nology in an outpatient context: a mixed methods trial evaluation. BMC Health Serv Res. 2026;26:182. doi:10.1186/s12913-025
aff-reported errors are distinct from independently measured note-error rates.

man A, James CA. Effective engagement with AI is the only path forward for clinician-educators. Acad Med. 2025;100(9 suppl
://pubmed.ncbi.nlm.nih.gov/40456164/ — Commentary on active engagement with AI while protecting foundational learning an
ting supervision.