ALLIANCE for ACADEMIC INTERNAL MEDICINE

Annual Review of the 2024 Medical Education Literature

Tina Mullick, MD, MS
Amy Kwon, MD
Elexis McBee, DO, MPH
Marina Mutter, MD, MS
Amy Weinstein, MD, MPH

Even Better Together

About our group

- Members of the Alliance for Academic Medicine (AAIM) – Clerkship Directors in Internal Medicine (CDIM) Survey and Scholarship Committee
- Annual systematic review of medical education literature, specifically undergraduate clinical training
- Select sample of studies that are broadly applicable to trainees at various levels
- We are all clinician-educators but not necessarily content experts in these areas
- Questions at the end of talk

Literature Review Search Strategy

("Clinical Clerkship"[MeSH Terms] OR "Curriculum"[MeSH Terms] OR "Education"[MeSH Terms] OR "education, medical"[MeSH Terms] OR "education, medical, undergraduate"[MeSH Terms] OR "hospitals, teaching"[MeSH Terms] OR "Learning"[MeSH Terms] OR "schools, medical"[MeSH Terms] OR "Teaching"[MeSH Terms] OR "clerkship*"[Title/Abstract] OR "curricul*"[Title/Abstract] OR "educat*"[Title/Abstract] OR "train*"[Title/Abstract] OR "measurement*"[Title/Abstract] OR "assess*"[Title/Abstract] OR "grade"[Title/Abstract] OR "grading"[Title/Abstract]) AND ("students, medical"[MeSH Terms] OR "student*"[Title/Abstract] OR "acting intern*"[Title/Abstract] OR "junior intern*"[Title/Abstract] OR “subintern” [Title/Abstract] OR “sub-intern” [Title/Abstract] OR "transition to residency"[Title/Abstract]) AND ("Acad Med"[Journal] OR "Adv Health Sci Educ Theory Pract"[Journal] OR "Can Med Educ J"[Journal] OR "Clin Teach"[Journal] OR "Educ Prim Care"[Journal] OR "Eval Health Prof"[Journal] OR "Health Aff"[Journal] OR "J Am Med Assoc"[Journal] OR "J Gen Intern Med"[Journal] OR "J Grad Med Educ"[Journal] OR "J Hosp Med"[Journal] OR "J Int Assoc Med Sci Educ"[Journal] OR "JAMA"[Journal] OR "Med Teach"[Journal] OR "Med Sci Educ"[Journal] OR "MedEdPORTAL"[Journal] OR "N Engl J Med"[Journal] OR "Teach Learn Med"[Journal]) **AND** (2024/1/1:2024/12/31[pdat])

- PubMed Search

Included:

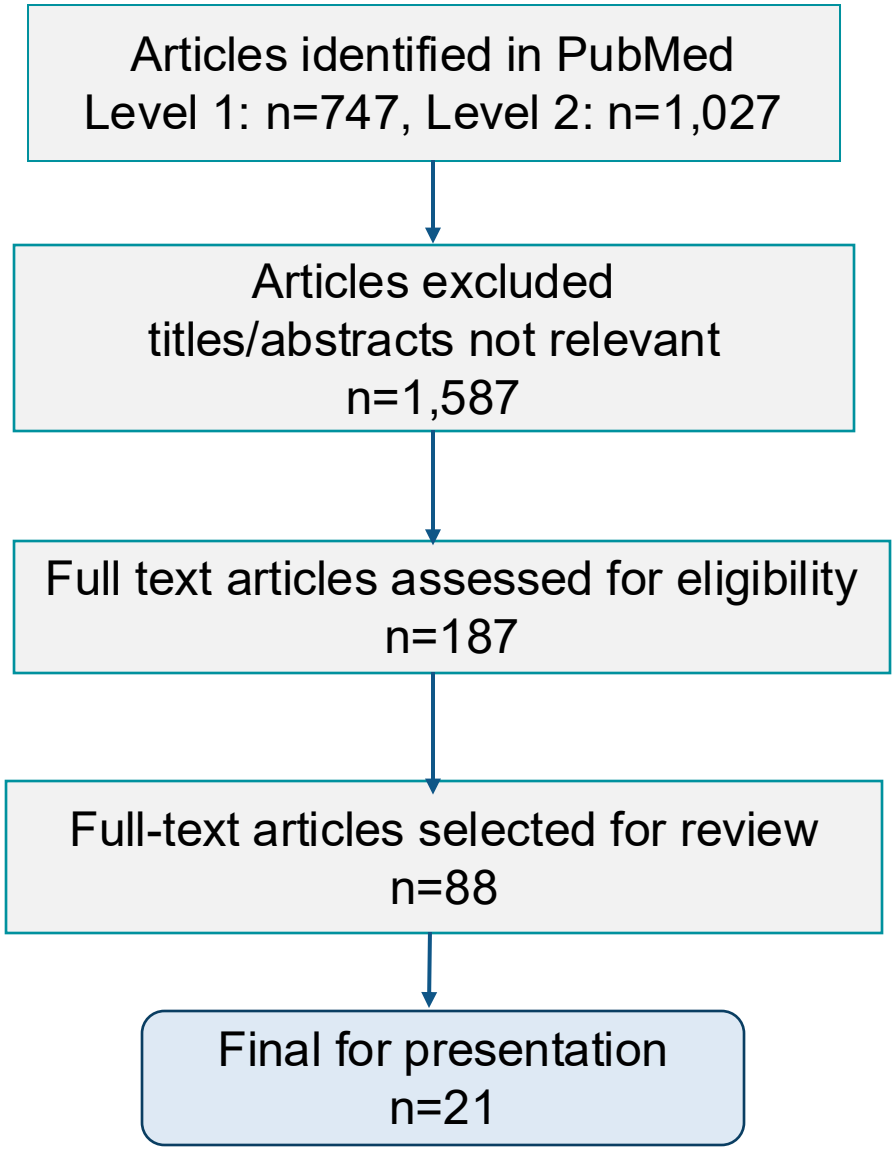
- UME articles
  - Level 1- no APC
  - Level 2 - APC
- Med Ed Journals
- Jan 1, 2024 to Dec 31, 2024

Excluded:

- Non-UME articles
- Commentaries
- Editorials
- Perspectives

Article Selection Process

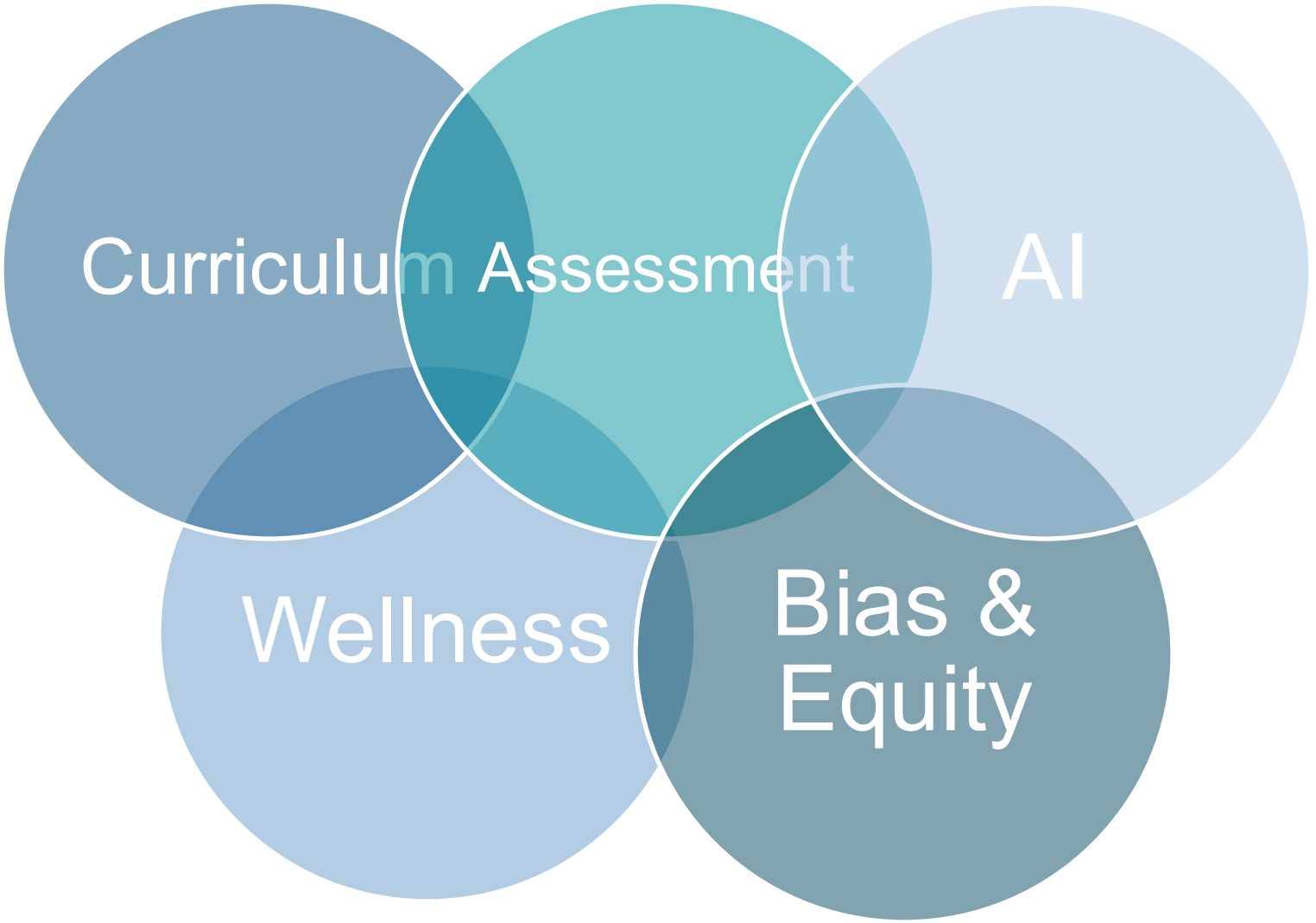
Katherine Downton, MSLIS
Librarian
University of Colorado Anschutz Medical Campus



Updated Scoring Rubric

Article:	Good (1)	Very good (2)	Excellent (3)
Rigorous study design (Consider large N, multiple institutions, results beyond student opinion, control group, randomization, validated survey instrument, power analysis; for qualitative - multiple coders, triangulation)			
Applicable/of interest to our CDIM audience			
Broad impact/importance			
Innovative/novel			
Total Points:			
Priority score: 1 – not important; 2 – may be; 3 - must include			
Comments:			

Core Themes



ALLIANCE for ACADEMIC INTERNAL MEDICINE

Student Assessment

Amy Weinstein, MD, MPH
Harvard Medical School

ALLIANCE for ACADEMIC INTERNAL MEDICINE

When you are asked to assess professionalism, what qualities do you assess?

DOI: 10.1111/tct.13830

RESEARCH ARTICLE

THE CLINICAL TEACHER

Student professionalism in clerkships

Lauren W. Cochran[1] | Lisa Auerbach[1] | Nicole K. Roberts[2]

Cochran. Clin Teach 2024

How can we better define, teach and assess student professionalism?

Background

- Professionalism can be challenging to define and assess
- Narratives tend to be vague
- Current literature focuses on breaches rather than milestones
- Students who struggle are more likely to struggle later

Cochran. Clin Teach 2024

Student Professionalism in Clerkships: Methods

- **Design:** Qualitative study
- **Setting:** Single 7-year BS/MD program in Harlem
- **Methods:**
  - 753 formative & summative core clerkship final narratives for 99 students were reviewed
  - Team developed a code, professionalism narratives were categorized & themes were identified

Cochran. Clin Teach 2024

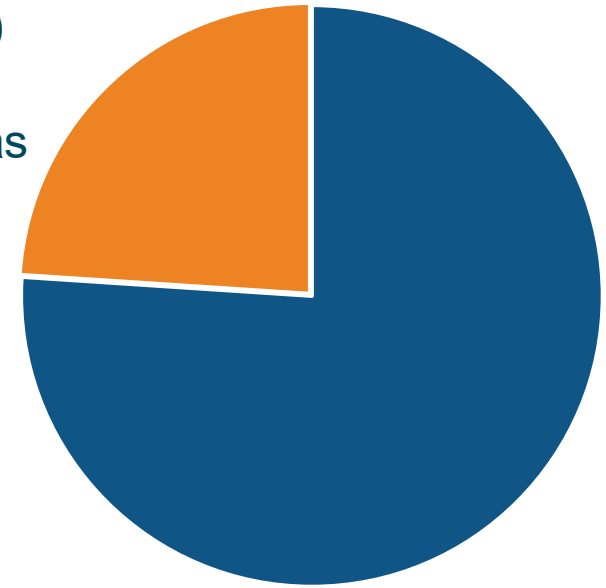
Student Professionalism in Clerkships: Results

Demographics:

- 63% women, 37% men
- 38% Asian, 31% Black/African American, 12% Latinx, 16% White and 2% unknown

Narratives (753)

Professionalism
23% (176)
95% were described as positive



LEVEL OF DETAIL

LOW:	MEDIUM:	HIGH:
75 (43%)	72 (41%)	29 (16%)

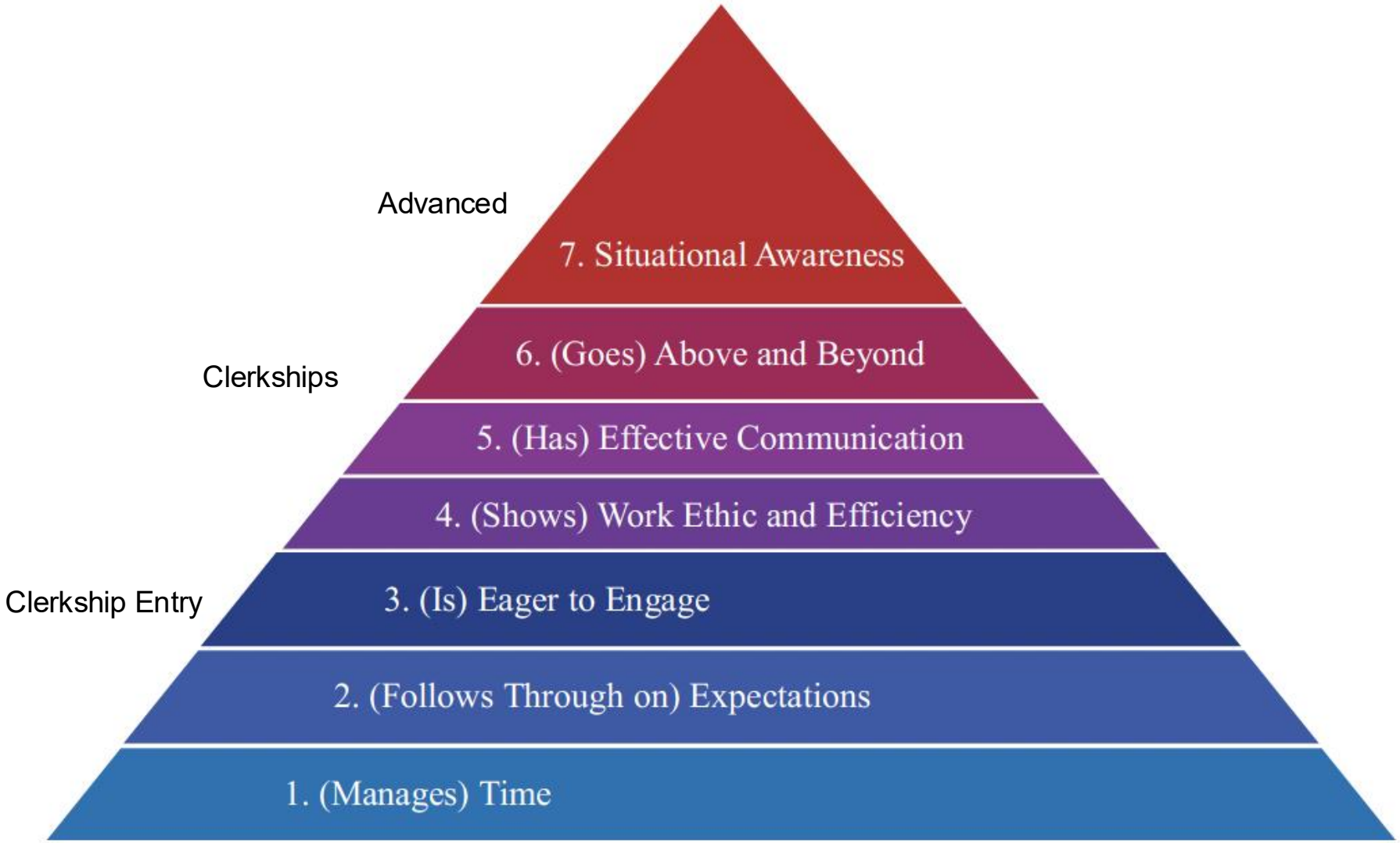
Cochran. Clin Teach 2024

Professionalism Assessment Themes and Sub-Themes

Theme	Sub-Theme
Manages Time	On time Comes early/stays late
Follows Through on Expectations	Prepared Takes ownership Completes assignments/requirements
Eager to Engage	Actively engaged Eager to try new activities
Work Ethic & Efficiency	Able to prioritize Uses downtime well
Effective Communication	Clinically Administratively
Goes Above & Beyond	Performs above expected level Innovates care systems
Situational Awareness	Interacts effectively Attuned to surroundings Enhances patient care

Cochran. Clin Teach 2024

Hierarchy of Professional Behaviors



Cochran. Clin Teach 2024

Student professionalism in clerkships: Limitations and take home points

Limitations

- Single new school, >40% URIM

A professionalism framework can define positive professionalism behaviors and support learning and assessment

- Students: supports self-directed learning, emphasizes growth mindset
- Faculty: supports faculty development, helps describe and evaluate professionalism
- Supports more specific formative and summative feedback

professionalism

noun | pro·fes·sion·al·ism | \prə-ˈfesh-nə-ˌli-zəm, -ˈfe-shə-nə-ˌli-\

Cochran. Clin Teach 2024

Moving Beyond the Dichotomous Assessment of Professionalism in the Internal Medicine Clerkship: Results of a National Survey of Clerkship Directors

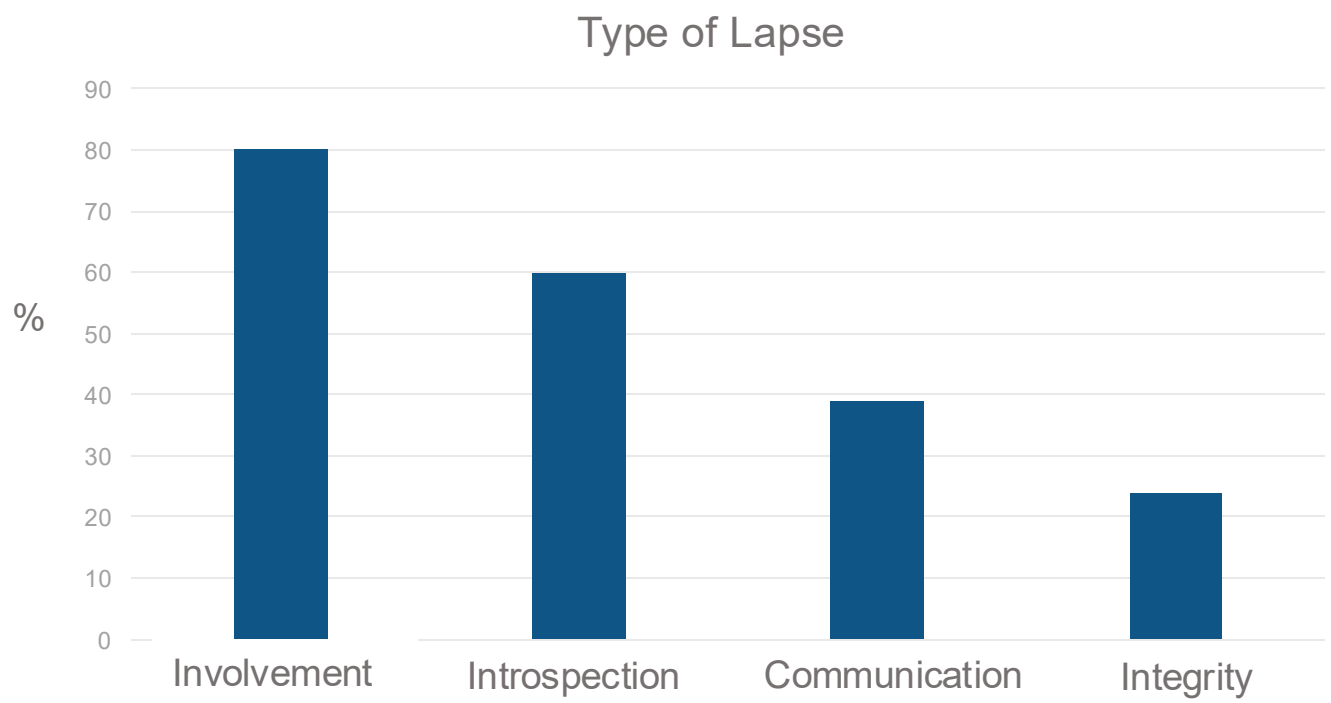
Mary W. Montgomery, MD, Elizabeth M. Petersen, MD, MPH, Amy R. Weinstein, MD, Camilla Curren, MD, Kathryn Hufmeyer, MD, Michael Kisielewski, MA, Edward Krupat, PhD, and Nora Y. Osman, MD

ACADEMIC MEDICINE | AAMC
Journal of the Association of American Medical Colleges

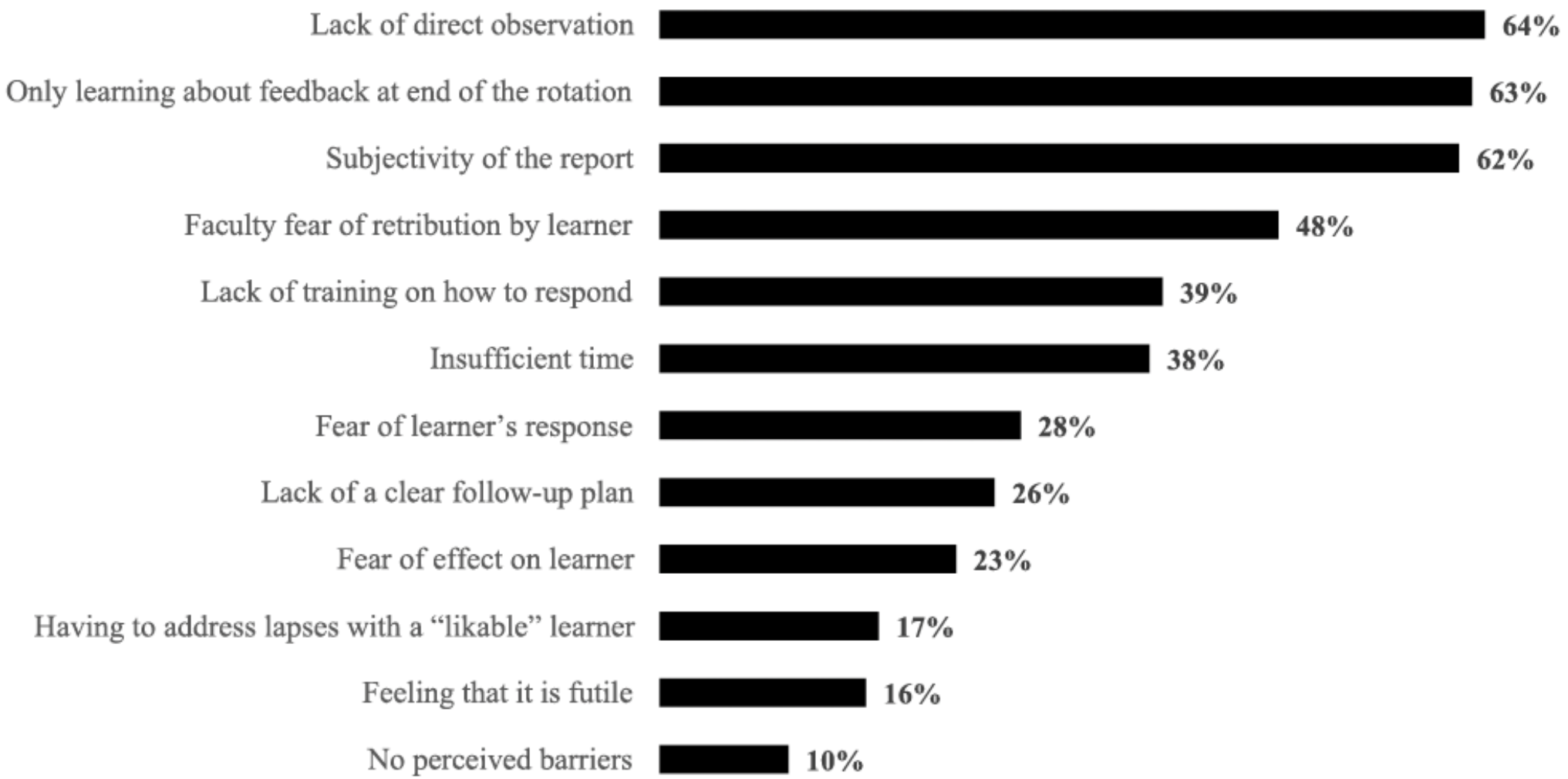
How do IM clerkships think about and assess professionalism?

2021 CDIM national survey examined how IM clerkship directors view professionalism

- 103 CDs (75.2%) responded
- 94% of clerkships formally evaluate professionalism
- 62% factor professionalism into clerkship grades



Barriers to assessing professionalism



How CDs perceive professionalism: Take home points

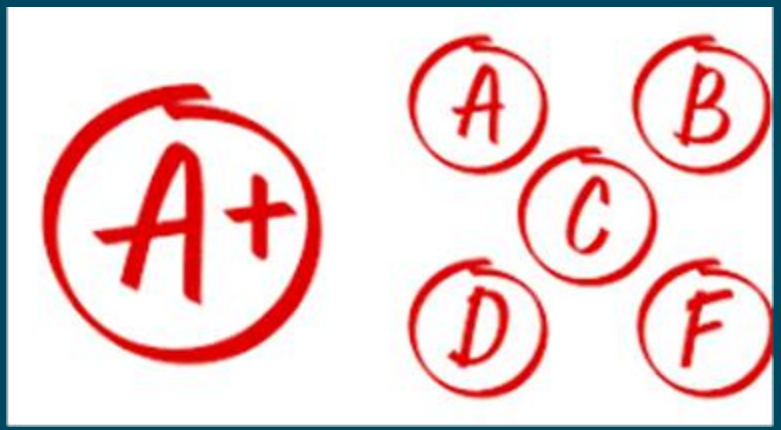
- Variety of professionalism issues demonstrate the complexity of assessment
- Most clerkships lack standard assessment and documentation practices
- Many barriers to identifying and addressing professionalism

- Current state focuses on lapse and remediation
- We should move towards a developmental model that nurtures growth

ALLIANCE for ACADEMIC INTERNAL MEDICINE

Bias and Equity in Medicine

Debaroti Tina Mullick, MD, MS
University of Tennessee Health Science Center
Memphis, TN



UT HSC

Even Better Together

What challenges and opportunities do you see in addressing bias and promoting equity within medical education?

ALLIANCE for ACADEMIC INTERNAL MEDICINE

Bias Reporting in the Clinical Learning Environment: A National Survey of Internal Medicine Clerkship Directors

Bruce L Henschen, Jennifer Pascoe, Michael Kisielewski, Nicholas S Duca, Valerie J Lang, Diane Levine, Harish Jasti, and Amy Blatt

Academic Medicine 2024; 99; 76-82

Bias Reporting in the Clinical Learning Environment: A National Survey of Internal Medicine Clerkship Directors

Background

- Bias exists in IM learning environment
- Unclear how often bias is identified
- How is bias addressed
- What are best practices for mitigation

Bias Reporting in the Clinical Learning Environment: A National Survey of Internal Medicine Clerkship Directors

Methods and Results

- CDIM survey 152 Medical schools 73% response
- 76/98 (78%) CDs reported at least 1 bias event
- Median 3 events
- Microaggressions 57%
- Race/Ethnicity 64%
- Gender 55%
- Less common gender identity, sexual orientation, accent, disability, religion, age, culture

Bias Reporting in the Clinical Learning Environment: A National Survey of Internal Medicine Clerkship Directors

Sources of Clerkship Bias Events	Events no (%)
Individual interactions with attending physician	51 (70%)
Individual interactions with resident	40 (55%)
Individual interactions with patient or family member	26 (36%)
Individual interactions with another medical student	9 (12%)
Individual interactions with another clinical member (ie nurse, SW, pt care tech)	7 (10%)
Bias in clerkship the physical environment	6 (8%)
Bias in clerkship grading policy	3 (4%)
Bias in opportunities given to different groups of students	3 (4%)
Other	2 (3%)

Bias Reporting in the Clinical Learning Environment: A National Survey of Internal Medicine Clerkship Directors

Mechanisms for identifying bias events	Events No (%)
End of clerkship evaluation	54 (56%)
Email to clerkship director	54 (56%)
Evaluation of an individual faculty member by a student	52 (54%)
Informal conversation with student	45 (46%)
Evaluation of an individual resident by a student	41 (42%)
AAMC graduate questionnaire	41 (42%)
Mid-clerkship feedback meeting	38 (39%)
Feedback from institutional reporting system	34 (35%)
Verbal report from resident	34 (35%)
Verbal report from faculty	32 (33%)
Clerkship exit interview	16 (17%)
Other	5 (5%)
None of the above	16 (17%)

Bias Reporting in the Clinical Learning Environment: A National Survey of Internal Medicine Clerkship Directors

Bias reports are at same level or increasing

Why?

- Heightened awareness of systemic racism
- Increased burnout leading to unprofessional behavior
- CDs more aware of learning climate, so students report more
- Bias still underreported and no clear best practice

Racial and Ethnic Bias in Letters of Recommendation in Academic Medicine: A Systematic Review

Saarang R Deshpande, Gina Lepore, Lily Weiland, and Jennifer R Kogan

Academic Medicine 2024; 99; 1032-1037

Racial and Ethnic Bias in Letters of Recommendation in Academic Medicine: A Systematic Review

Background

Letters of recommendation (LORs) are key components for residency and fellowship applications

- Prior systematic review found significant gender bias in LORs
- Standardized LORs (SLORs) more objective with less bias?
- Disparities in medical education for under-represented in medicine (URiM) students (less AOA, fewer honors)

Racial and Ethnic Bias in Letters of Recommendation in Academic Medicine: A Systematic Review

Methods

Systematic review of all published studies from inception through July 2023

- LORs/SLORs
- Racial, ethnic, URiM bias
- Physicians/trainees in medicine

Included 23 studies, 19,012 applicants, 41,925 LORs (82.6% residency)

Racial and Ethnic Bias in Letters of Recommendation in Academic Medicine: A Systematic Review

Results: Linguistics

- Fewer agentic terms in URiM *(some studies no difference)*
  - “skill, strong, lead, confidence, active, work ethic”
- More communal terms in Black/Latinx *(some studies no difference)*
  - "agreeable, warm, considerate"
- More grindstone terms in URiM *(some studies no difference)*
  - "dedicated, hardworking, organized"
- More standout terms in white applicants
  - “exceptional, best, outstanding”
- Emotive terms for anger more in URiM
- More “doubt raisers” in URiM (hedging, veiled criticism, faint praise)

Racial and Ethnic Bias in Letters of Recommendation in Academic Medicine: A Systematic Review

Results:

Topic differences

- Less discussion of surgical skills in URiM

SLORs

- URiM had fewer linguistic differences in SLORs compared to LORs
- URiM had lower global assessment scores/ratings in some studies

Racial and Ethnic Bias in Letters of Recommendation in Academic Medicine: A Systematic Review

Discussion

- Bias against URiM in LORs in academic medicine
  - Heterogeneity, small numbers of URiM applicants
  - Skewed towards procedural specialties
  - Many studies with no difference, size of effect difficult to ascertain
- SLORs may reduce (but not eliminate) racial and ethnic bias
- Goal for no bias at all! Or to develop a tool for recognizing bias in letters. Possibly through AI?

“We Need A Seismic Shift”: Disabled Student Perspectives on Disability Inclusion in US Medical Education

Neera R Jain, Erene Stergiopoulos, Amy Addams, Christopher J Moreland, and Lisa M Meeks

Academic Medicine, November 2024, Vol 99, NO 11, 1221-1232

“We Need A Seismic Shift”: Disabled Student Perspectives on Disability Inclusion in US Medical Education

Background

- Students with disabilities encounter inequities in medical education
  - Poor training conditions
  - Structural, cultural, and climate associated barriers
  - Persistent disability-related stigma
- Many students do not disclose disability or request accommodations
- If access needs not met, higher risk of poor outcomes
  - Longer time to graduation, lower board scores
  - Higher burnout rates, increased rates of depression
- Culture of ableism within medical education

“We Need A Seismic Shift”: Disabled Student Perspectives on Disability Inclusion in US Medical Education

Methods

AAMC Year 2 questionnaire in 2019 and 2020

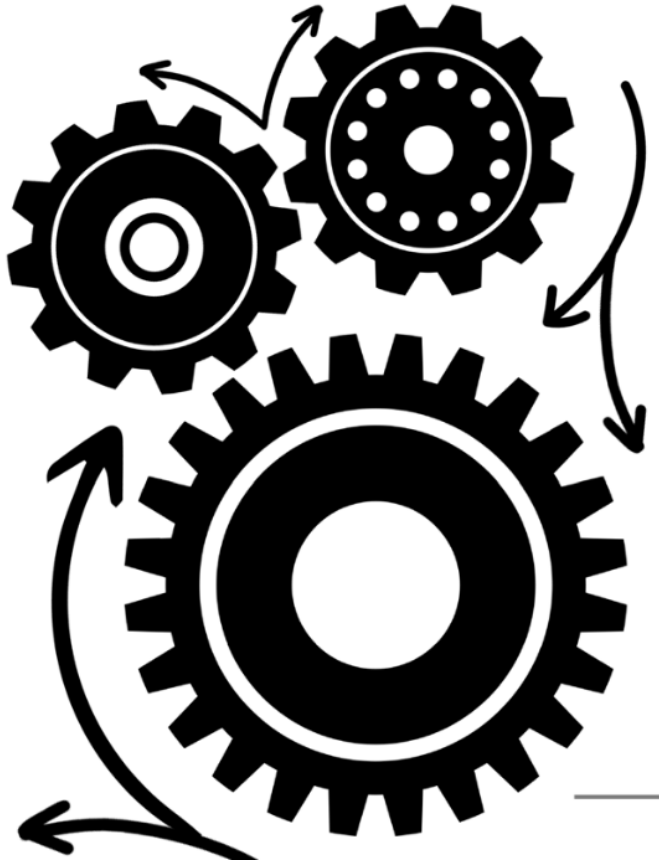
- 61% response rate overall (27k students), 9% self-reported disability
- Thematic analysis of narrative responses to below question:
  - “Use the space below if you would like to share anything about your experiences regarding medical school and disability”

Results

- 674 responses, (27% of those with disabilities)
- **De-identified data to protect anonymity

“We Need A Seismic Shift”: Disabled Student Perspectives on Disability Inclusion in US Medical Education

Results Part 1: Mapping the System



PROGRAM STRUCTURE

Foundational to student experiences—determines whether they need to interact through processes and people to seek access.

PROCESS

If students need to seek adjustment to the program structure, the processes to access those adjustments shape their experience—positively and negatively.

PEOPLE

Relationships with faculty, clinical educators, administrators, and fellow students shaped student interactions with disability inclusion processes and their experience of medical education. People influenced students’ experiences of process and could be influential in shifting the structure and culture.

CULTURE

Informed by and built through experiences with structure, process, people. Also informed by wider society and historical experiences.

“We Need A Seismic Shift”: Disabled Student Perspectives on Disability Inclusion in US Medical Education

Results Part 1: Mapping the System

Program structure: how curricula are taught, workload/schedule, physical and electronic environment

- If barriers existed -> all learners out to learn in the same prescribed ways
- Inaccessible conditions added to students’ labor
- Favorable elements – built-in flexibility

Process: process to broker accessibility, support well-being

- Need for clear, efficient, transparent…not “othering” processes
- Burdensome -> significant time and advocacy

“We Need A Seismic Shift”: Disabled Student Perspectives on Disability Inclusion in US Medical Education

Results Part 1: Mapping the System

People: relationships with faculty, administrators, students

Quality and attitudes mattered – can make conditions difficult or foster seamless engagement

Students perceived a lack of power against harmful interactions

Culture:

Perception of ableist culture

“We Need A Seismic Shift”: Disabled Student Perspectives on Disability Inclusion in US Medical Education

“I consistently feel like I am a fish being judged by my ability to fly - needless to say, it is extremely frustrating, exhausting, and makes me constantly question my worth as a student and whether I belong here” [program structure]

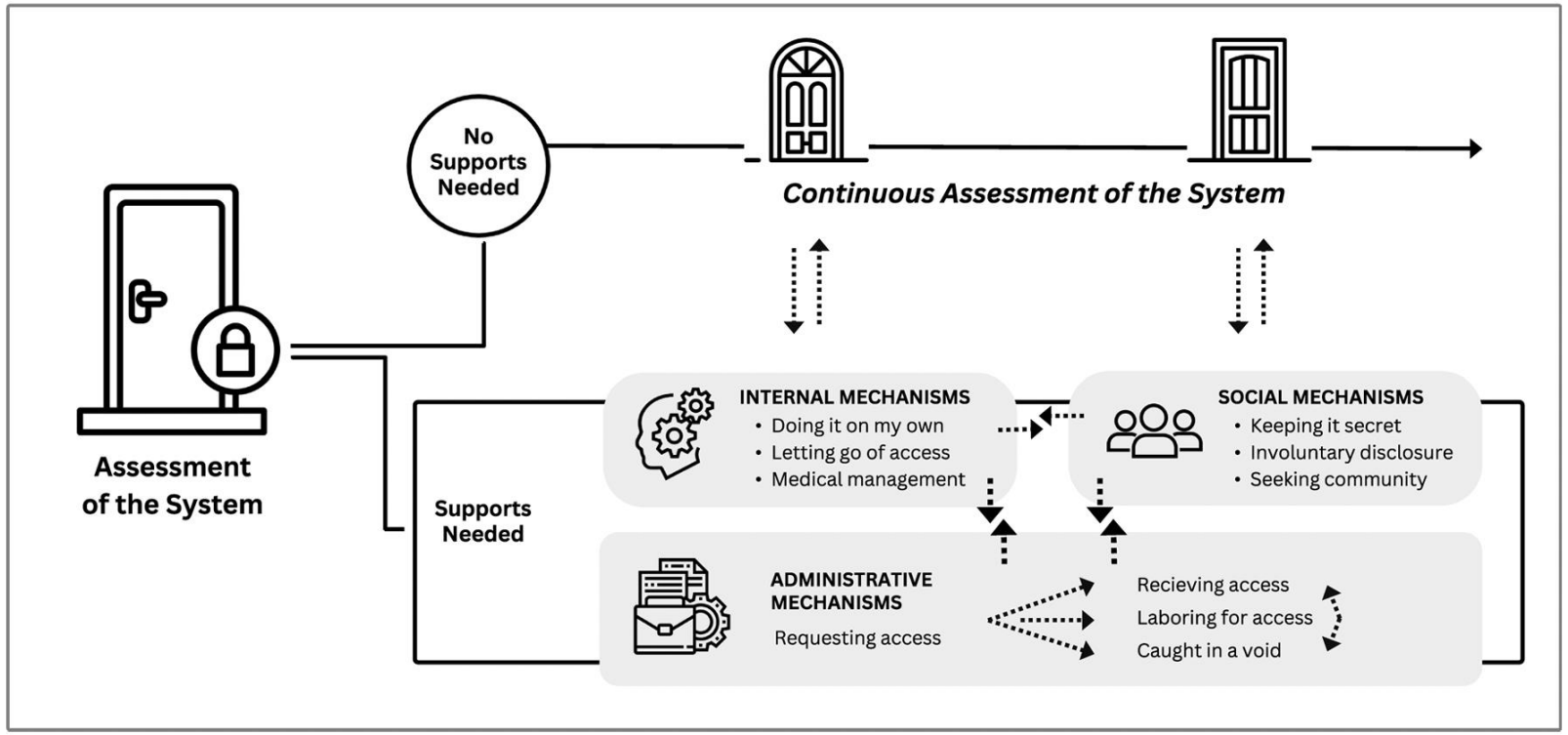
"Incredibly difficult to request accommodation originally when my school was requesting I go through the global university resources..." [process]

"I've faced a lot of discrimination and harassment from administrators and faculty regarding my psychological disabilities..." [people]

“We Need A Seismic Shift”: Disabled Student Perspectives on Disability Inclusion in US Medical Education

Results Part 2: Navigating the System

Students navigated the system through administrative, social and internal mechanisms



“We Need A Seismic Shift”: Disabled Student Perspectives on Disability Inclusion in US Medical Education

Results Part 2: Navigating the System

"I have had to do a lot of self-advocating [for accommodations] and I frequently feel as if I were a nuisance" [administrative]

"I know the types of accommodations offered to students with ADHD. I don't feel these accommodations would benefit me enough to make up for the stigma of identifying myself as a person with ADHD" [social]

"I haven't really found that I have needed any form of accommodation during school. I have struggled with my diagnosis since I was about 9 years old and have found very decent coping strategies as well as medication that works for me" [internal]

“We Need A Seismic Shift”: Disabled Student Perspectives on Disability Inclusion in US Medical Education

Discussion:

- Analysis of largest known dataset of disabled medical students’ perspectives
- Highly divergent experiences, initial hurdle was program structure
  - Some students had overwhelmingly positive experiences
- Disability was something to be managed, remains highly stigmatized
- *Limitations*:
  - No demographic data to contextualize experiences
  - Second year students, have not entered clinical training

“We Need A Seismic Shift”: Disabled Student Perspectives on Disability Inclusion in US Medical Education

“Medicine is still made by and for abled people and on the basis of an abled physician”

“I am hoping that people with power in medicine (you!) will actually recognize the need for a seismic shift in how we approach disabilities”

ALLIANCE for ACADEMIC INTERNAL MEDICINE

Wellness & Curriculum

Marina Mutter, MD, MS
University of Colorado School of Medicine

"When You're in It, It Feels Like It's Everything": Medical Students' Experience of Failure and Remediation in the United States and the Netherlands

Lynnea M. Mills, MD, Terese Stenfors, PhD, Melissa Duffy, MA, PhD, John Q. Young, MD, PhD, Christy Boscardin, PhD, MA, Olle ten Cate, PhD, and Patricia S. O'Sullivan, EdD, MS

Grush, K. , Christensen, W. , Lockspeiser, T. & Adams, J. (2025). Secondary Traumatic Stress in Medical Students During Clinical Clerkships. *Academic Medicine, 100* (3), 325-330. doi: 10.1097/ACM.0000000000005940.

Overview

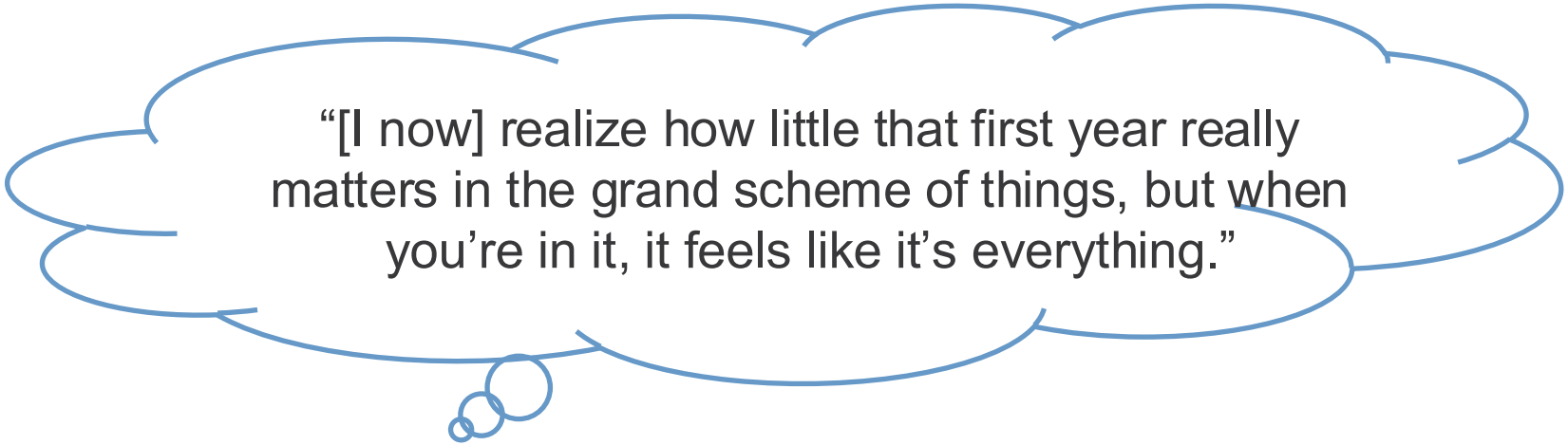
- **Aim:** Explore medical students' *emotional* experience of failure and/or remediation

- **Setting/participants**: Virtual interviews with 9 current medical students in the US and 5 in the Netherlands from July to September 2022 who had not met expectations on at least one medical school assignment
  - US: Typical 4-year curriculum
  - Netherlands: Time-variable curriculum (most 6.5+ years)

- **Data analysis:** Codebook thematic analysis of interview transcripts

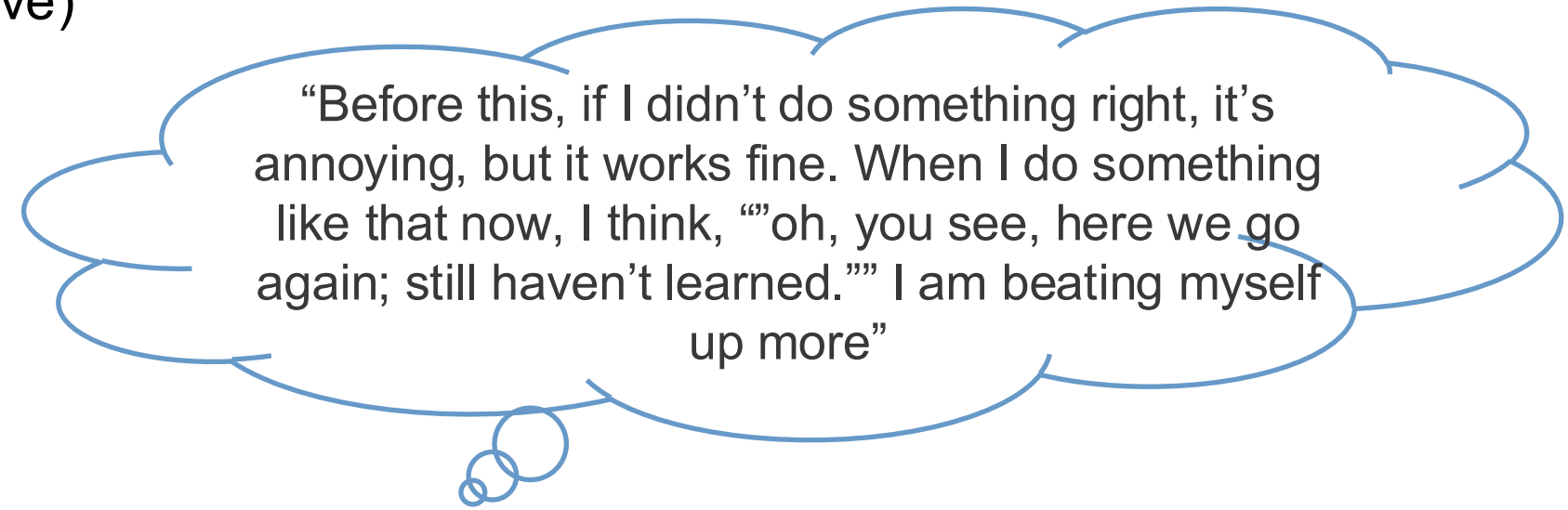
Results

- At US schools, all participants failed exams; at Dutch schools 4/5 failed a clerkship
  - Most had successfully completed remediation
- All felt that the failure and remediation was a significant event in their education
- Shame was pervasive, with self-doubt, resentment and blame also common
  - Many assumed that they were unique in having failed



Results

- With time and additional perspective, negative emotions softened (but often did not fully resolve)



- Worry and stress surrounding career effects also present but differed between countries (less in the Netherlands, in a time-variable educational system)

Conclusions

- Negative emotions are common, improved (but did not fully resolve) over time; impact on professional identify formation
  - Greater need to normalize failure, build community of same-experience peers and/or role models?

- *Limitations:* No US medical students failed a clerkship/clinical experience, all current students, unclear if any students remediated a whole year (or longer), potential bias towards students comfortable with sharing experience

ALLIANCE for ACADEMIC INTERNAL MEDICINE

Artificial Intelligence (AI)

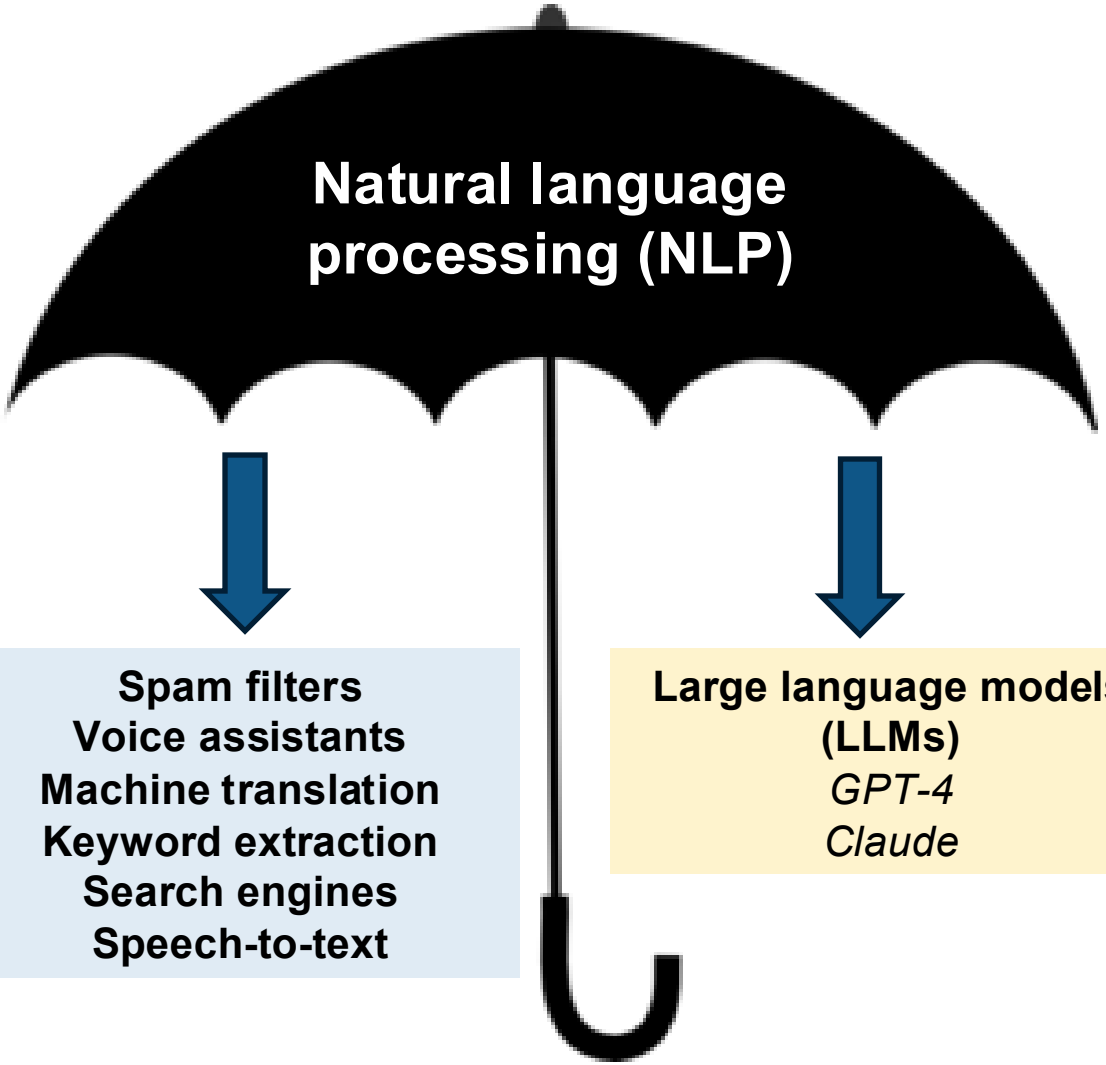
Amy Kwon, MD
Oregon Health & Science University

What are some ways that generative AI has been used at your institution in the teaching realm?

Are there areas where you are excited about AI in your own work?

Worries or fears?

Some terms and concepts to know



Allows machines to understand, interpret, and generate human language

A *type* of NLP model — specifically, very large neural networks trained on vast text datasets — specializing in understanding and generating human-like text **(generative AI)**

Comparing Scoring Consistency of Large Language Models with Faculty for Formative Assessments in Medical Education

Radhika Sreedhar, MD, MS[1], *Linda Chang, PharmD, MPH*[1], *Ananya Gangopadhyaya, MD*[1], *Peggy Woziwodzki Shiels, MD*[1], *Julie Loza, MD*[1], *Euna Chi, MD*[1], *Elizabeth Gabel, MD*[1], *and Yoon Soo Park, PhD*[1]

[1]University of Illinois College of Medicine, Chicago, IL, USA

JGIM. 2025 Jan;40(1):127-134.

Problem: Narrative written feedback on formative written assignments are a significant time investment for faculty.

Question: Can large language models (LLMs) provide accurate and consistent narrative feedback for formative assessments in the undergraduate medical education (UME) setting?

Study design (Sreedhar et al)

- **Design:** Single institution, multi-site retrospective cross-sectional study of pre-clinical students' critical appraisal assignments during one academic year
- **Setting:** University of Illinois College of Medicine (UICOM) at Chicago, Peoria, and Rockford campuses, academic year 2022-2023
- **Dataset**: 111 discrete pre-clinical critical appraisal assignments that did not meet at least 1 grading rubric criterion
- **LLM:** ChatGPT 3.5

2022-2023
Assignments submitted
Reviewed by faculty

111/600 assignments hand-selected for GPT review

Prompt with de-identified assignments submitted to GPT

Developed with sample set of 10 assignments and training prompt

Faculty assessment v ChatGPT scoring evaluated

Assignments did not meet ≥1 grading rubric criterion
Intentionally distributed across rubric

EPA 7	Does not meet expectations
Self-identified learning need (SDL) Question 1	Learning needs related to the patient is **not** clearly described
PICO (ASK) Questions 2 and 3	Clinical question is missing PICO components without explanation
Resource (ACQUIRE) Question 4	Resource is **not** a primary source
Strengths and weaknesses (APPRAISE) Question 5	Student does not describe the strengths and weaknesses of the article
Strengths and weaknesses (APPRAISE) Question 5	Student does not describe the strengths and weaknesses of the article
Answer (APPLY) Question 6	Student's summary of how the study addresses the patient's problem is not explained

No statistically significant differences between faculty and ChatGPT assessments, but issues with application and context

Table 2 Descriptive Statistics and Percent Exact Agreement Between ChatGPT and Faculty Scoring (No. = 111)

checklist item	Chat GPT scoring No. (%) of students who met criteria	Faculty scoring No. (%) of students who met criteria	% exact agreement between ChatGPT and faculty scores	*P*-value
Is the learning need related to the patient clearly described?	99 (89%)	88 (79%)	74%	0.70
Does the clinical question have all the PICO components?	106 (95%)	93 (84%)	81%	0.81
Is the citation that was used primary literature?	104 (94%)	63 (57%)	60%	0.12
Does the student describe the strengths and weaknesses of the article cited?	88 (79%)	81 (73%)	63%	0.91
Does the student summary explain how the study cited addresses the patient's problem?	86 (77%)	66 (59%)	59%	0.39

Examples of discordant feedback between faculty and ChatGPT

ChatGPT feedback	Faculty feedback
Overall, the student did a good job of critically appraising the article and summarizing the information found in a way that could be presented to a preceptor. The student's decision to use this therapy for their patient is justified by the information presented in the article	

Take-home points (Sreedhar et al)

Question: *Can large language models (LLMs) provide accurate and consistent narrative feedback for formative assessments?*

- **Answer: Not yet for high-stakes situations, but possibly soon.**
- Reasonable to consider AI assistance in evaluating **formative assessments**, but this is **not yet ready for high-stakes prime time**
- With ongoing interim developments and refinements in LLMs, this development could be sooner rather than later

Barriers/Limitations: *Privacy concerns, misinterpretation or ignoring of student content, inherent bias of LLM, potential cost barrier of LLM usage*

Next steps: *More specific prompt engineering*

Enhancing self-directed learning with custom GPT AI facilitation among medical students: A randomized controlled trial

Wang Shalong[a], Zuo Yi[b], Zou Bin[c], Liu Ganglei[a], Zhou Jinyu[a], Zheng Yanwen[a], Zhang Zequn[a], Yuan Lianwen[a] and Ren Feng[a]

[a]Department of General Surgery, The Second Xiangya Hospital of Central South University, Changsha, China; [b]School of Information Technology and Management, Hunan University of Finance and Economics, Changsha, China; [c]Department of General Surgery, The Affiliated Changsha Central Hospital Hengyang Medical School, University of South China, Changsha, China

Med Teach. 2024 Oct 19:1-8.

Problem: Self-directed learning is an essential skill for physicians to acquire for professional success.

Question: Can LLMs help medical trainees develop self-directed learning skills?

Study design (Shalong et al)

- **Design:** Single center, open-label randomized controlled trial for 14 weeks (12-week intervention period followed by 2-week follow up period)
- **Setting:** Second Xiangya Hospital of Central South University (China), November 2023-February 2024
- **Participants**: 103 undergraduate and postgraduate students with access to ChatGPT, specializing in clinical medicine at Second Xiangya Hospital. *Exclusion: LLM usage for learning assistance >1X weekly for prior 2 months*
- **LLM:** ChatGPT (with custom GPT LearnGuide)

103 Randomized

51 Intervention group

ChatGPT + custom GPT + utilization course
2 hour intro session

52 Control group

ChatGPT only

Baseline

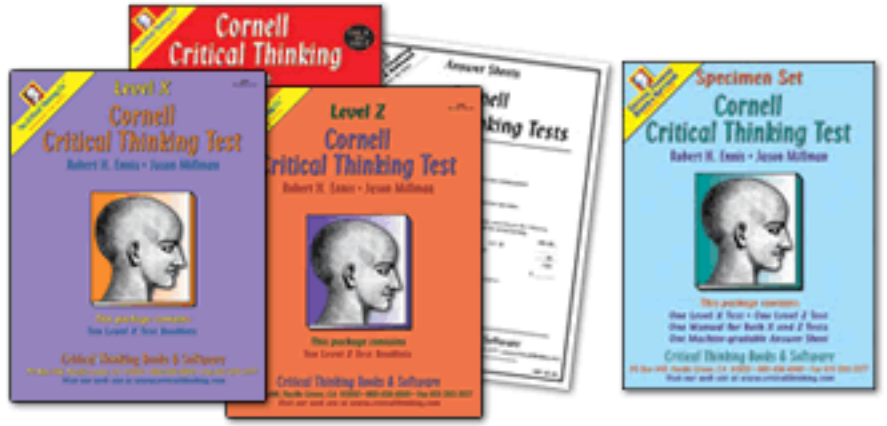
6 weeks

12 weeks

2-weeks post

12-week course with integrated weekly problem-based learning assignments

2-week washout period



Self-Directed Learning Scale (SDLS)
Cornell Critical Thinking Test (CCTT)

Statistically significant, sustained improvements seen in critical thinking at the end of the intervention period

Table 2. Baseline SDLS, CCTS, and GFS scores of participants.

Characteristic/baseline scores	Control group ($n = 52$)	Intervention group ($n = 51$)	Intervention vs control, mean difference (95% CI)	*P*-valuea
Baseline SDLS Score (median [IQR]) (range, 10-50)	27.0 (24.0 to 30.5)	28.0 (26.5 to 30.0)	1.3 (-0.2 to 2.8)	0.10
Baseline CCTS score (median [IQR]) (range, 0-52)	25.0 (22.0 to 28.5)	24.0 (21.5 to 28.0)	1.1 (-3.1 to 0.9)	0.23

Table 3. Outcome variables at intervention and follow-up.

Outcome variable	Median (IQR) Control group ($n = 51$)	Intervention group ($n = 50$)	Intervention vs control, mean difference (95% CI)	*P*-value[a]
SDLS score, (range, 10-50)				
6 wk	28.0 (23.5 to 31.0)	31.0 (26.0 to 37.8)	3.1 (0.0 to 6.2)	0.05
12 wk	29.0 (24.0 to 35.0)	33.5 (26.3 to 42.0)	4.2 (0.8 to 7.5)	0.01
14 wk	28.4 (24.3 to 32.3)	34.4 (30.4 to 39.2)	5.9 (3.2 to 8.5)	<.001
CCTT score, (range, 0-52)				
6 wk	24.0 (20.0 to 29.5)	25.0 (21.3 to 31.8)	1.3 (-1.6 to 4.2)	0.27
12 wk	27.0 (22.5 to 29.0)	33.5 (29.0 to 39.5)	7.1 (4.5 to 9.7)	<0.001
14 wk	26.0 (23.0 to 30.0)	33.0 (29.0 to 36.0)	6.0 (4.0 to 8.0)	<0.001

Take-home points (Shalong et al)

Question: *Can LLMs help medical trainees develop self-directed learning skills?*

- **Answer: Possibly, with appropriate training and guidance.**
- Access to LLMs is not enough: Appropriate training, required practice in utilization, and custom GPTs may have all contributed to durable improvement in self-directed learning metrics

Additional barriers/Limitations: *Generalized (rather than medicine-specific) SDL metrics, selection bias toward those with technology access, barriers to building custom GPTs, effective integration of custom GPTs into existing educational frameworks without inappropriate overuse*

Making Use of Natural Language Processing to Better Understand Medical Students' Self-Assessment of Clinical Skills

Laurah Turner, PhD, Danielle E. Weber, MD, Sally A. Santen, MD, PhD, Amy L. Olex, PhD, Pamela Baker, PhD, Seth Overla, MS, Derek Shu, Anne Randolph, MHA, and Matt Kelleher, MD, MEd

Acad Med. 2024;99:285-289.

Problem: Reviewing large-volume reflective self-assessment narratives is a time-consuming task.

Question: Can natural language processing (NLP) tools identify useful themes in large-volume free-text narratives?

Study design (Turner et al)

- **Design:** Single center, cross-sectional study across 2 distinct academic years
- **Setting:** University of Cincinnati College of Medicine, during academic years 2019-2020 and 2021-2022
- **Dataset**: 2,224 reflective self-assessments from 344 pre-clinical medical students after standardized clinical encounters
- **NLP:** TopEx (proprietary, developed by Virginia Commonwealth University)

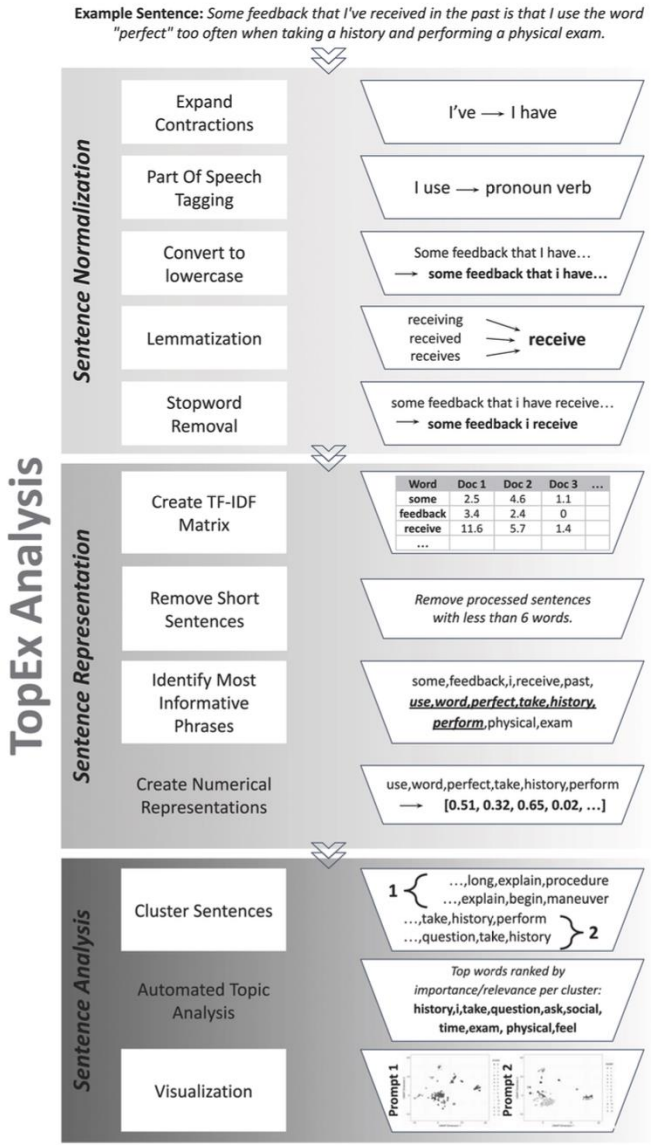
Standardized clinical encounter

Reflective prompts

1. *Comment on 1-2 aspects of your performance that surprised you.*
2. *Write a specific thing to improve in future patient interactions.*

TopEx analysis

- *Sentence normalization*
- *Sentence representation*
- *Sentence analysis*



Cluster analysis
Individual authors

Theme extraction
Consensus review by authors + NLP expert

Theme	Frequency
Prompt 1: Please comment on 1–2 aspects of your performance that surprised you	
Organization and approach to asking questions	215
Analysis of history taking (sexual history, social history, etc.)	125
Guidance of patient and overall execution of the physical exam	100
Confidence in explaining exam maneuvers	97
Positive reflections on abilities	87
Appearance/perception of body language	83
Talking/speech style and nonverbal behaviors	76
Eye contact	46
Filler words	41
Time spent looking at notes and/or screens	23

Prompt 2: Write one specific thing you could do to improve in future patient interactions	
Taking sensitive histories	171
Improving the efficiencies and maneuvers of the physical exam and explaining the physical exam to the patient	114
Changing the speed of speech	89
Asking more open-ended questions	87
Using filler words such as "um," "okay," etc.	57
Listening to standardized patient	47
Making eye contact	38
Explaining things to the patient during the physical exam and avoiding jargon	38

n = 2,224 free-text narratives

Potential applications: Normalizing growth areas, guiding curricular improvement, identifying deficiencies on an individual or group level

Take-home points (Turner et al)

Question: *Can natural language processing (NLP) of large-volume free-text narratives help identify areas for growth?*

- **Answer: Yes, it can help identify themes from large amounts of text.**
- **Potential applications:** Assessing global themes regarding clinical skills (and deficiencies) across student cohort
  - *Normalized feedback*
  - *Practical assessment revisions*
  - *Curricular improvement*

Barriers: *Unfamiliarity with NLP and lack of user-friendly NLP applications, need for manual review/confirmation, technical assistance with NLP expertise*

Final take-home points

- A **professionalism framework** can help both trainees and faculty define and target specific behaviors, and encourage a developmental rather than binary approach
- **Bias (especially racial, gender, disability) persists** in the clinical environment and beyond, and may be addressed by standardized approaches
- **Failure/remediation has durable negative emotional impacts** among US medical trainees
- **Generative AI** has many potentially useful applications in content creation, trainee coaching, and data analysis, but requires **careful editing/interpretation** of output and **proper training** on appropriate usage

Thank you!

ALLIANCE for ACADEMIC INTERNAL MEDICINE

Questions?

Tina Mullick, MD, MS
dborsche@uthsc.edu

Amy Kwon, MD
kwonam@ohsu.edu

Marina Mutter, MD, MS
MARINA.MUTTER@cuanschutz.edu

Amy Weinstein, MD, MPH
aweinste@bidmc.harvard.edu