

Contemporary* approaches to assessment of learners**

**and potentially innovative*
***Allergy and Immunology fellows*

Katherine L. Tuttle, MD
Associate Program Director, University of Rochester
February 23rd, 2024



UNIVERSITY of ROCHESTER

1

Disclosures

Funding [related]: 2021 AAAAI Educator Development Award

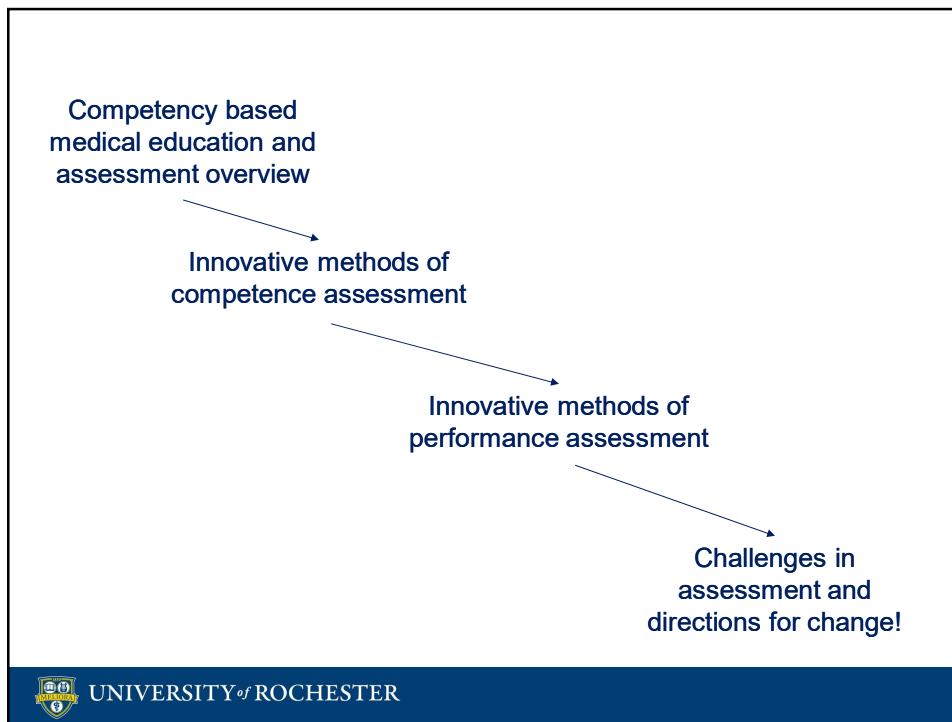
Funding [not related]: DBV Technologies

1. I am not anti-multiple choice test!
2. I frequently participate in affinity bias, and is something I work on!



UNIVERSITY of ROCHESTER

2



3

Assessment meaning and history in medical education


Latin root: *assidere to sit beside*

3 main eras:

1. Assessment as measurement
2. Assessment as judgement
3. **Assessment as a system [programmatic assessment]**

Competency based medical education [CBME] relies on effective, reliable, and equitable assessment.

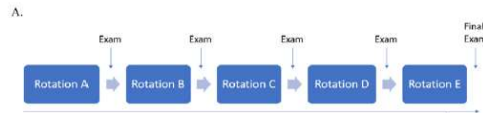
What is the next era?

 UNIVERSITY of ROCHESTER

Schuwirth and van der Vleuten Adv Health Sci Educ Theory Pract 2020

4

CBME centers on continuous assessment



5

Assessment types: formative and summative

Formative assessment: intended to inform feedback to guide a learner's development rather than provide a final evaluation.

Assessment **for** learning

Summative assessment: intended to measure current status of competency

Assessment **of** learning



6

CBME formative assessment overarching principles

Frequent
Perceived as 'low stakes'
Often but not always workplace based
Individual learning trajectory
Rich with trainee feedback

Evolution towards a **co-production learning cycle**



UNIVERSITY of ROCHESTER

G.B. Lee and A.M. Chiu / Ann Allergy Asthma Immunol 128 (2022) 256–262
Holmboe et al. Acad Med. 2023;98:S37–S49.

7

Miller's pyramid, *chronic spontaneous urticaria*



Figure 3. Methods of assessment.¹⁶ OSCE, objective structured clinical examination.



UNIVERSITY of ROCHESTER

G.B. Lee and A.M. Chiu / Ann Allergy Asthma Immunol 128 (2022) 256–262

8

Join by Web PollEv.com/katetuttle428

Join by Text Send [katetuttle428](https://PollEv.com/katetuttle428) to [22333](https://PollEv.com/katetuttle428)



 UNIVERSITY of ROCHESTER

9

Innovative methods of competence
assessment
[2 evidence based examples]

 UNIVERSITY of ROCHESTER

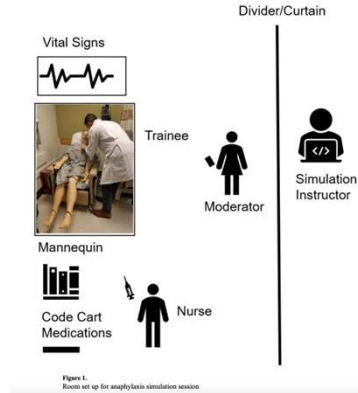
10

Simulation as assessment of competence

37 events, 19 AI/ trainees
Pre questionnaire 1 week before

5 simulation anaphylaxis cases:

1. Anaphylaxis after IT
2. Isolated hypotension as presentation of anaphylaxis in an adult
3. Anaphylaxis in an infant
4. Anaphylaxis during a pediatric food challenge
5. Anaphylaxis in an adult on a beta blocker



Post questionnaire 1-2 weeks after



UNIVERSITY of ROCHESTER

Barmettler et al. J Allergy Clin Immunol Pract. 2020 ; 8(10): 3616–3618.
Mawhirt SL, Fonacier L, Aquino M. Ann Allergy Asthma Immunol Off Publ Am Coll Allergy Asthma Immunol. 2019 5;122(5):513–21.

11

Simulation: effective assessment of competence

Table 1.

Anaphylaxis knowledge and confidence results pre- and post-simulation.

Knowledge Question	% Correct Pre-Simulation (n=37 [*])	% Correct Post-Simulation (n=23)	p-value [†]
Identification of situations that make anaphylaxis likely	89	96	p<0.001
Risk factors for poor outcomes	57	70	p=0.008
Epinephrine concentration	84	96	p<0.001
Epinephrine dosing in adults/pediatrics	97	100	p<0.001
First line treatment for patients on beta-blockers	87	91	p<0.001
Initial anaphylaxis management steps	35	48	p=0.84
Why early administration of epinephrine essential	100	100	N/A
	%Moderately Confident/Very Confident Pre-Simulation (n=37 [*])	%Moderately Confident/Very Confident Post-Simulation (n=23)	
Confidence regarding anaphylaxis recognition/management	76	100	p=0.04

* There were 37 AI trainee participation events from 19 individuals

† McNemar's test



UNIVERSITY of ROCHESTER

Barmettler et al. J Allergy Clin Immunol Pract. 2020 ; 8(10): 3616–3618.
Mawhirt SL, Fonacier L, Aquino M. Ann Allergy Asthma Immunol Off Publ Am Coll Allergy Asthma Immunol. 2019 5;122(5):513–21.

12

Simulation: can involve multisource feedback, identified challenge

Multisource feedback [Dr. Ward's presentation in Jan meeting highlighted this!]

- simulation can/should involve teams, specifically nursing
- assess communication with health care team

Faculty involvement is time intensive in increasing formative assessments in CBME, particularly in simulation

Simulation can be expensive.

Despite training, inter-reviewer reliability can be variable



UNIVERSITY of ROCHESTER

13

Automated assessment models: artificial intelligence and machine learning

Anaphylaxis simulation from *Managing Emergencies in Pediatric Anesthesia*

Model input: 52 video data set, 54% fail, 46% pass

Model: bidirectional transformer encoder

Classes: Resnet, GloVe common crawl

Model output: pass/fail

Table 1
Incorporated Input Features from High-Fidelity Anaphylaxis Simulation in Model Development for Automated Assessment of Trainees

Evaluation model	Model input features
Model 1	1. Transcription text 2. Raw audio 3. Raw video 4. Vital signs
Model 2	1. Raw audio 2. Raw video 3. Vital signs
Model 3	1. Transcription text 2. Raw audio 3. Vital signs
Model 4	1. Transcription text 2. Raw audio 3. Raw video
Model 5	1. Transcription text 2. Raw video 3. Vital signs



UNIVERSITY of ROCHESTER

Siddiqui et al. Acad Med. 2023;98:1274-1277.

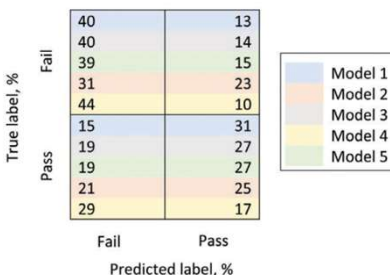
14

Automated assessment models require additional research

Table 2

Accuracy, Recall, Precision, and F1 Score for Models 1-5 for Automated Assessment of Trainees in High-Fidelity Simulation of an Anaphylaxis Scenario

Outcome	Model 1	Model 2	Model 3	Model 4	Model 5
Accuracy	0.71	0.56	0.67	0.62	0.65



UNIVERSITY of ROCHESTER

Siddiqui et al et al. Acad Med. 2023;98:1274-1277.

15

Automated assessment models currently only partially address CBME assessment goals

Anaphylaxis isn't pass fail! Can clinical nuance be captured?

Growth opportunities:

Vital sign interpretation

Clinical reasoning, epinephrine use,

Challenging to evaluate other core competencies using automated assessment models [professionalism].



UNIVERSITY of ROCHESTER

16

Innovative methods of performance
assessment

[2]



UNIVERSITY of ROCHESTER

17

Consider bundling competence and
performance assessments together!



UNIVERSITY of ROCHESTER

18

Direct observation as performance assessment

REACT tool: Rapid Evaluation Assessment of Clinical Reasoning

REACT score: minimum 3, maximum 15

41 fourth year medical students at UVA trialed in clinical simulations of urgent clinical situations

Hypotension
Chest Pain
AMS
Hypoxemia

7 raters

Cronbach's alpha: 0.86
Weighted kappa: 0.56
[moderate agreement]

Essential Tasks		Assessment of Behaviors				
Learner Function	Specified Task	Instructions: Circle the box corresponding to the learner's observed behavior				
Collecting • Data gathering - recognition of urgent or emergent clinical scenario	• Collect/Report history and exam data in history/oriented manner • Recognize patient and disease specific factors as potential etiologies of decompensation • Recognize severity of clinical problem and contextual signs of urgency or emergency	<input type="checkbox"/> Non focused history and exam <input type="checkbox"/> Included extraneous information <input type="checkbox"/> Missed key findings <input type="checkbox"/> Did not recognize contextual clues of urgency	<input type="checkbox"/> History and exam reflect potential diagnosis <input type="checkbox"/> Limited recognition of urgent contextual clues <input type="checkbox"/> Included limited pertinent positive and negative findings	<input type="checkbox"/> Logical history and exam for potential diagnosis <input type="checkbox"/> Questions assessed likelihood of specific diagnosis <input type="checkbox"/> Full recognition of urgency contextual clues <input type="checkbox"/> Prioritization on pertinent positive and negative findings		
Interpreting • Diagnostic reasoning - differential diagnosis	• Generate prioritized differential diagnosis of most likely, less likely, unlikely in urgent clinical situations	<input type="checkbox"/> Differential diagnosis missing likely or "can't miss" diagnoses in urgent situations <input type="checkbox"/> Includes inappropriate diagnoses	<input type="checkbox"/> Differential diagnosis included likely and "can't miss" diagnoses in urgent situations but missed key diagnoses <input type="checkbox"/> Inappropriate rank-order of diagnoses	<input type="checkbox"/> Accurately ranked differential diagnosis including key, likely, and "can't miss" diagnoses in urgent situations <input type="checkbox"/> Prioritized urgent diagnoses appropriately		
Managing • Management reasoning - initial management option selection - response to dynamic information	• Direct evaluation and treatment towards high priority diagnoses • Initiate management to patient with urgent decompensation • Recognize need to evaluate patient care	<input type="checkbox"/> Directed evaluation and treatment toward unlikely/unimportant diagnoses <input type="checkbox"/> Did not evaluate or treat most likely urgent diagnosis <input type="checkbox"/> Did not evaluate for response to initial management plan	<input type="checkbox"/> Major focus of evaluation and treatment on likely and urgent diagnoses <input type="checkbox"/> Included non-essential testing <input type="checkbox"/> Evaluated for response to initial management plan	<input type="checkbox"/> Efficiently directed management towards most likely and urgent diagnoses <input type="checkbox"/> Deferred tests directed towards less likely or less important diagnoses <input type="checkbox"/> Evaluated for response to initial interventions		
Communicating • Patient-centered care - communicate care plan and goals of care - achieve a shared mental model	• Communicate with health care team members according to role and responsibility to improve task efficiency • Maintain focus and necessary communication with patient or surrogate decision-maker in urgent clinical situations	<input type="checkbox"/> Did not engage health care team members <input type="checkbox"/> Did not communicate care plan with patient/family <input type="checkbox"/> Did not clarify patient's goals of care	<input type="checkbox"/> Limited engagement of health care team members to improve task efficiency <input type="checkbox"/> Limited effort to communicate with patient/family but used medical jargon or failed to ensure understanding <input type="checkbox"/> Limited clarification of patient's goals of care	<input type="checkbox"/> Fully engaged health care team members according to role and responsibility to improve task efficiency in urgent situation <input type="checkbox"/> Effectively communicated care plan with patient/family and ensured understanding <input type="checkbox"/> Effective clarification of patient's goals of care		
Reflecting • Metacognition - reflection of urgent or emergent clinical scenarios	• Demonstrate the ability to think about one's own thinking (metacognition) in urgent clinical situations • Mitigate cognitive tendencies or emotional/situational factors influencing clinical decision making	<input type="checkbox"/> Unaware of cognitive tendencies or emotional/situational factors that may have influenced decision-making <input type="checkbox"/> Unable to mitigate cognitive tendencies or emotional/situational factors	<input type="checkbox"/> Limited awareness of cognitive tendencies or emotional/situational factors that may have influenced decision-making <input type="checkbox"/> Limited mitigation of cognitive tendencies or emotional/situational factors	<input type="checkbox"/> Full awareness of cognitive tendency or emotional/situational factors that may influence decision-making <input type="checkbox"/> Effectively mitigates reactions to urgent clinical situations		

Figure 1 Rapid Evaluation Assessment of Clinical Reasoning Tool (REACT).



UNIVERSITY of ROCHESTER

Siddiqui et al. Acad Med. 2023;98:1274-1277.

19

Assessment meaning and history in medical education

Latin root: *assidere to sit beside*

3 main eras:

1. Assessment as measurement
2. Assessment as judgement
3. **Assessment as a system [programmatically assessment]**

What is the next era?



UNIVERSITY of ROCHESTER

Schuwrth and van der Vleuten Adv Health Sci Educ Theory Pract 2020

20

Performance assessment innovations

What is the next era of assessment?

Why does performance matter?
high quality and equitable care for patients.

How does one assess this at the trainee level?

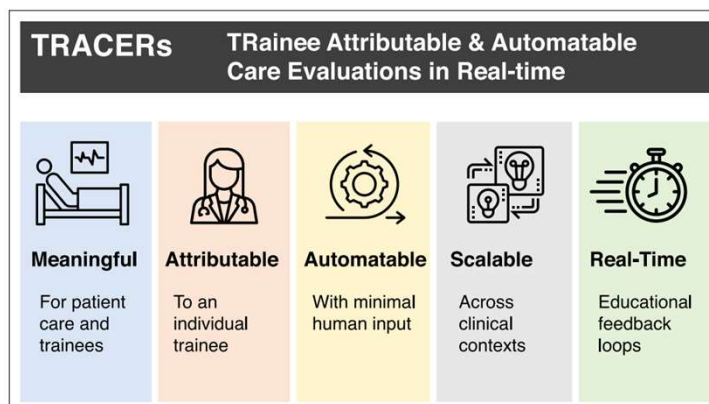


UNIVERSITY of ROCHESTER

Schumacher DJ et al. Acad Med 2023

21

TRACERs: framework for assessing clinical trainee performance to inform system level outcomes



UNIVERSITY of ROCHESTER

Burk-Rafel et al. Perspectives on Medical Education. 2023;12(1):149-159

22

TRACER example

	CPM	TRACER
	PERCENTAGE OF OUTPATIENTS WITH HEMOGLOBIN A1C (HbA1C) >9.0%	ORDER LONG-ACTING INSULIN FOR INPATIENTS WITH TYPE 2 DM AT RISK FOR HYPERGLYCEMIA
Meaningful (for patient care)	Important outcome for patients	Guideline-directed behavior known to improve glycemic control
Meaningful (for trainees)	Does not provide trainees with feedback on which aspects of their practices affect HbA1c	Indicates trainee understanding of diabetes and insulin types
Attributable	HbA1c levels are product of individual, patient, and system-level factors and may not be predominantly attributable to a single trainee	Reasonably attributable to trainee placing medication orders
Automatable	HbA1c automatically calculated from electronic health record panel data	Medication orders recorded by and retrievable from electronic health record
Scalable	Employable across clinical settings	Applicable in diverse institutional settings
Real-time	Relatively sparse measure that requires a minimum of 3 months to accrue updates and may not drive just-in-time feedback	Amenable to near real-time feedback

Data returned to trainee and fellowship team from clinical informaticists, feedback near immediate



UNIVERSITY of ROCHESTER

Burk-Rafel et al. Perspectives on Medical Education. 2023;12(1):149-159

23

Challenges with linking education to patient care in Allergy/Immunology

Specific diagnoses can be rare.

TRACERs

Order icatibant for outpatients with hereditary angioedema

Check tryptase for new patient visit outpatients with a new diagnosis of venom hypersensitivity with hypotension

Propose desensitization for outpatients with type I drug hypersensitivity with limited treatment options



UNIVERSITY of ROCHESTER

24

What is burdensome about assessment?

What makes it difficult to 'sit beside'?



UNIVERSITY of ROCHESTER

25

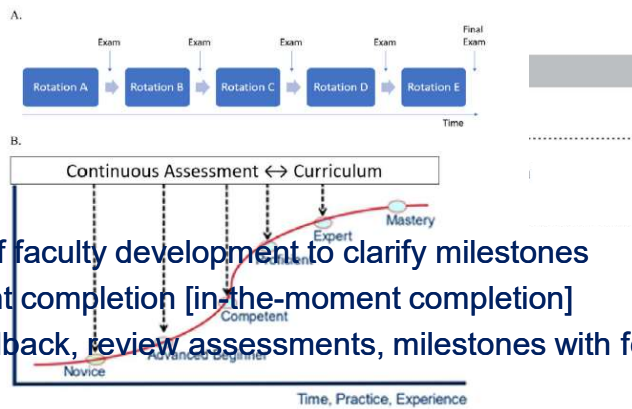
Significant assessment burden themes

Table 1

Overview

Theme

Disparate n
processes in
Challenges
processes (t



Evolution of faculty development to clarify milestones

Assessment completion [in-the-moment completion]

Fellow feedback, review assessments, milestones with fellow and faculty

Dedicated clinical time for faculty for direct clinical observation



UNIVERSITY of ROCHESTER

Szulewski et al. Acad Med. 2023;98:1261-1267

26

What about from the trainee perspective?

What makes it difficult for the trainees to 'sit beside' with us?



UNIVERSITY of ROCHESTER

27

Case study

K.L. a medical student [she/her], is on her pediatrics rotations in third year. K.L. identifies as URiM. After the team's resident notices KL conversing in Haitian Creole, she assigns KL 2 Creole speaking patients and asks her to interpret with other patients.

During attending rounds, K.L. demonstrates excitability and is constantly describing to the residents how much she loves pediatrics. She stutters during the case presentations. The resident interrupts her to get rounds 'going'. The attending frequently mispronounces her name. .



UNIVERSITY of ROCHESTER

28

Case study

The attending and resident spend most of the non clinical conversations discussing their alma mater, a small liberal arts college in Western Massachusetts.

K.L receives feedback from the resident that her feelings are not genuine. Faculty member calls her 'too assertive'.

Following the feedback, KL appears less engaged.



UNIVERSITY of ROCHESTER

29

More formative, lower stakes, assessments? Bias in frontline clinical assessments

1. **Contextual equity** consider the context of the environment in which learners are assessed.
2. **Intrinsic equity** consider the assessment tools
3. **Instrumental equity** considers how we communicate and use our assessments

Involve DEI experts, health equity researchers



UNIVERSITY of ROCHESTER

Onumah et al. Acad Med. 2023;98:S57-S63.

30

Contextual equity strategies

1. Build an environment that promotes equity
 - Role model use of trainees preferred name, use correct pronouns
 - Avoid affinity bias! Establish rapport with learners by asking about their values and passions [Ask: what are your non-medical interests? Not 'where did you go to college?']
2. Identify extra pressure the learning environment puts on some learners
3. Practice self reflection regarding personal bias



31

Instrumental equity strategies

1. Provide feedback to support growth
 - Foster a growth mindset: 'we all have areas for improvement and need feedback to grow'
 - Give specific feedback on observed behavior

K.L, you seem very excited, can you tell me more why and how you feel?

2. Revise narrative language to minimize bias
 - Use competency-based descriptors rather than personality descriptors

K.L. described the pathophysiology of asthma in exquisite detail, demonstrated compassion to family during asthma exacerbation, and identified barriers to care



32

Actionable steps forward for us, to help us 'sit beside'
as program directors!



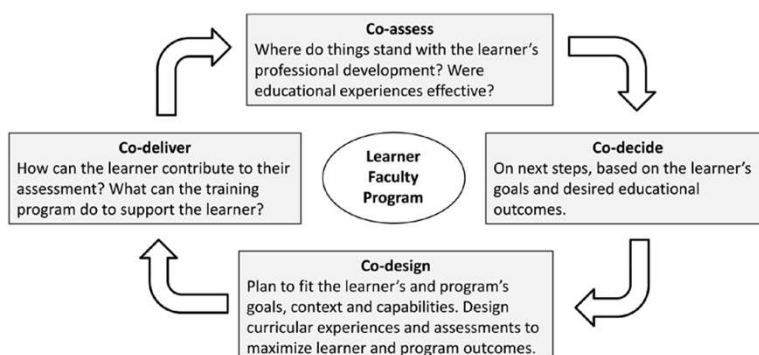
UNIVERSITY of ROCHESTER

33

Coproduction learning is a cycle, can reduce bias, allow for trainee feedback

A

Co-production Learning Cycle



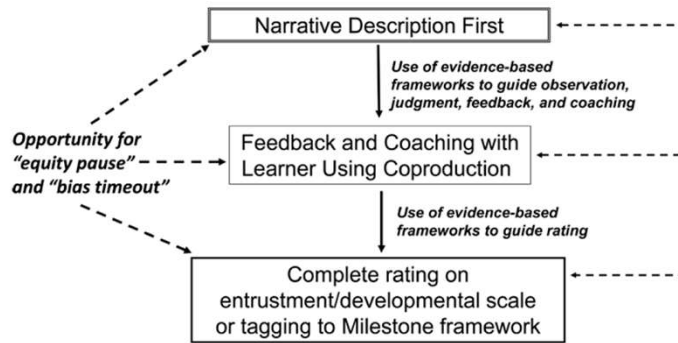
UNIVERSITY of ROCHESTER

Holmboe et al. Acad Med. 2023;98:S37-S49.

34

Contemporary and innovative assessment practice as a program

Rethinking the Assessment Process



UNIVERSITY of ROCHESTER

Holmboe et al. Acad Med. 2023;98:S37-S49.

35

Tool for diagnosing gaps within your assessment system

Components	Complete each item to discover strengths and uncover potential weaknesses within your competency-based assessment system.	In place	
		Yes	No
Fairness Policies and procedures meet standards for equity and equality. ^{1,4}	Program Requirements: Faculty and learners can easily access written program requirements and assessment policies.	<input type="checkbox"/>	<input type="checkbox"/>
	Due Process: Written policies exist in clearly posted areas to guide learners who wish to appeal performance decisions.	<input type="checkbox"/>	<input type="checkbox"/>
	Equity: Learners within a program have comparable experiences (didactic sessions, patient mix/volume, etc.) and feedback opportunities.	<input type="checkbox"/>	<input type="checkbox"/>
	Bias: Program administrators or leaders periodically monitor relationships between performance decisions and learners' characteristics (gender, race/ethnicity, verbal fluency, etc.) for potential bias.	<input type="checkbox"/>	<input type="checkbox"/>
Reliability Mechanisms exist that contribute to consistent performance decisions. ^{1,2,4}	Faculty Development: Assessors have a shared understanding of performance expectations for different levels of learners across different contexts.	<input type="checkbox"/>	<input type="checkbox"/>
	Adequate Sampling: Sufficient assessment evidence exists to document learners' performance using multiple methods (exams, clinical performance ratings, simulations, etc.) from different assessors (faculty, peers, patients, etc.) across contexts (didactic sessions, clinical settings, research environments, etc.).	<input type="checkbox"/>	<input type="checkbox"/>
Validity Qualified assessors make accurate, defensible decisions about learner competence. ^{1,3}	Alignment: Curricular experiences and assessment evidence align closely with performance expectations stated in competencies, milestones, and entrustable professional activities (if in place).	<input type="checkbox"/>	<input type="checkbox"/>
	Qualified Assessors: Trained assessors provide specific feedback to document learners' competence within particular settings.	<input type="checkbox"/>	<input type="checkbox"/>
	Authentic Work: Assessment evidence documents learners' progress and performance within actual work-based settings.	<input type="checkbox"/>	<input type="checkbox"/>
Effects Nurtures reflective practice. ^{1,3}	Quality Evidence: Assessment evidence provides meaningful, accurate, and timely documentation of learners' competence.	<input type="checkbox"/>	<input type="checkbox"/>
	Educational Effects: Processes within the assessment system encourage learners to reflect on performance, to design and implement learning plans, to solicit feedback from others for ongoing improvement, and to monitor their progress and achievement.	<input type="checkbox"/>	<input type="checkbox"/>
Outcomes Intended and unintended consequences monitored. ^{1,2,4}	Acceptability: Faculty, learners, regulatory agencies, and employers view assessment decisions and effects as credible.	<input type="checkbox"/>	<input type="checkbox"/>
	External Measures: Program administrators or leaders monitor learners' performance (licensure exams, residency placements, etc.) for adverse or unanticipated outcomes possibly linked to assessment system.	<input type="checkbox"/>	<input type="checkbox"/>



UNIVERSITY of ROCHESTER

Bierer, Pien et al. Acad Med 2018;93:512

36

Thank you!

I am deeply humbled to be invited.

Dr. Theresa Bingemann
Program Directors Assembly [Dr. Khoury and Dr.
Sheikh]

Dr. Castells
Dr. Wickner
Dr. Hong
Dr. Schneider



UNIVERSITY of ROCHESTER