

Supplemental digital content for Smith S, Kogan JR, Berman NB, Dell MS, Brock DM, Robins LS. The development and preliminary validation of a rubric to assess medical students' written summary statements in virtual patient cases. Acad Med.

Supplemental Digital Appendix 1

Questions on a Reflection-on-Action Survey Completed by Raters Using an Assessment Rubric for Evaluating Medical Students' Summary Statements in Virtual Patient Cases

Q1: How would you describe your method of assigning the argument score? (Be as descriptive as possible about the general process you used.)

Q2: What do you think were the factors that most influenced your approach to assigning argument ratings?

Q3: Were there different factors in play for any of the cases? (Explain.)

Supplemental Digital Appendix 2

Correlation Between the Components of an Assessment Rubric for Evaluating Medical Students' Summary Statements, Across All Four Raters, By Individual Virtual Patient Case

Case and component	Factual accuracy	Appropriate narrowing of the differential diagnosis	Transformation of information	Use of semantic qualifiers	Global rating
Pediatrics case					
Factual accuracy	1.000				
Appropriate narrowing of the differential diagnosis	0.231	1.000			
Transformation of information	0.189	0.476 ^a	1.000		
Use of semantic qualifiers	-0.032	0.339	0.454 ^a	1.000	
Global rating	0.290	0.672 ^b	0.602 ^b	0.439 ^a	1.000
Family medicine case					
Factual accuracy	1.000				
Appropriate narrowing of the differential diagnosis	0.349 ^a	1.000			
Transformation of information	0.189	0.440 ^a	1.000		
Use of semantic qualifiers	0.213	0.094	0.251	1.000	
Global rating	0.393	0.829 ^b	0.663 ^b	0.271	1.000
Internal medicine case					
Factual accuracy	1.000				
Appropriate narrowing of the differential diagnosis	0.387	1.000			
Transformation of information	0.051	0.603 ^b	1.000		
Use of semantic qualifiers	0.285	0.382	0.433 ^a	1.000	
Global rating	0.332	0.894 ^b	0.659 ^b	0.346	1.000

^a $P < .01$ by Spearman's rho calculation.

^b $P < .001$ by Spearman's rho calculation.

Supplemental Digital Appendix 3

Results of a Content Analysis of Reflection-on-Action Comments Regarding Reasons for Assigning a Global Rating During Response Process Exploration

Scoring decision and themes	Example
Assigning the global rating	
Assess presence of key findings	There need to be enough key features (pertinent positives and negatives included). (Rater1)
Score individual items then score argument	I completed the individual scoring approach, then assigned a global score. (Rater3)
Most influential factor	
Accuracy	Accuracy was the most important issue. (Rater 3)
Organization	Focused on key findings, exclusion of extraneous information/detail, and organizing it into an argument. (Rater 4)
Most negative impact on global rating	
Inaccuracy	...misleads reader, e.g. inaccuracy ("no SOB" for patient with P.E.) or no ability to narrow ddx at all (e.g. refusing to walk - no ability to differentiate pain from weakness from imbalance, let alone differentiate between causes of leg pain or, more specifically, hip pain). (Rater 2)
Case-related factors	
Case is two-pronged	The DVT/PT case really has 2 diagnoses – DVT and PE. The other cases are more a single diagnosis. (Rater 2)
Case outside expertise	Doing cases outside your own area of expertise or knowledge base is more difficult. (Rater 4)