Development and Validation of a Tool to Evaluate the Evolution of Clinical Reasoning in Trauma Using Virtual Patients

David Fleiszer, MD,[*,†,1] Michael L. Hoover, PhD,[‡,2] Nancy Posel, PhD,[*,§] Tarek Razek, MD,[*,‖,3] and Simon Bergman, MD[*,¶,#,4]

[*]Faculty of Medicine, McGill University, Montreal, Quebec, Canada; [†]Departments of Surgery and Oncology, McGill University Health Center, Montreal, Quebec, Canada; [‡]Department of Educational and Counselling Psychology, McGill University, Montreal, Quebec, Canada; [§]McGill University Health Center, Montreal, Quebec, Canada; [‖]Montreal General Hospital, Montreal, Quebec, Canada; [¶]Department of Surgery, Jewish General Hospital, McGill University, Montreal, Quebec, Canada; and [#]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

CONTEXT: Undergraduate medical students at a large academic trauma center are required to manage a series of online virtual trauma patients as a mandatory exercise during their surgical rotation.

PURPOSE: Clinical reasoning during undergraduate medical education can be difficult to assess. The purpose of the study was to determine whether we could use components of the students' virtual patient management to measure changes in their clinical reasoning over the course of the clerkship year. In order to accomplish this, we decided to determine if the use of scoring rubrics could change the traditional subjective assessment to a more objective evaluation.

BASIC PROCEDURES: Two groups of students, one at the beginning of clerkship (Juniors) and one at the end of clerkship (Seniors), were chosen. Each group was given the same virtual patient case, a clinical scenario based on the Advanced Trauma Life Support (ATLS) Primary Trauma Survey, which had to be completed during their trauma rotation. The learner was required to make several key patient management choices based on their clinical reasoning, which would take them along different routes through the case. At the end of the case they had to create a summary report akin to sign-off. These summaries were graded independently by two domain "Experts" using a traditional subjective surgical approach to assessment and by two "Non-Experts" using two internally validated scoring rubrics. One rubric assessed procedural or domain knowledge (Procedural Rubric), while the other rubric highlighted semantic qualifiers (Semantic Rubric). Each of the rubrics was designed to reflect established components of clinical reasoning. Student's *t*-tests were used to compare the rubric scores for the two groups and Cohen's *d* was used to determine effect size. Kendall's τ was used to compare the difference between the two groups based on the "Expert's" subjective assessment. Inter-rater reliability (IRR) was determined using Cronbach's alpha.

MAIN FINDINGS: The Seniors did better than the Juniors with respect to "Procedural" issues but not for "Semantic" issues using the rubrics as assessed by the "Non-Experts". The average Procedural rubric score for the Senior group was 59% ± 13% while for the junior group, it was 51% ± 12% ($t_{(80)}$= 2.715; p = 0.008; Cohen's d = 1.53). The average Semantic rubric score for the Senior group was 31% ± 15% while for the Junior group, it was 28% ± 14% ($t_{(80)}$ = 1.010; p = .316, ns). There was no statistical difference in the marks given to the Senior versus Junior groups by the "Experts" (Kendall's τ = 0.182, p = 0.07).

Correspondence: Inquiries to Nancy Posel, Faculty of Medicine, McGill University, Suite 301, 4920 de Maisonneuve, Montreal, Que., Canada H3Z 1N1; e-mail: nancy.posel@mcgill.ca

[1] Royal Victoria Hospital, 1001 avenue Décarie, Room D2-7115, Montréal, QC. Canada H4A 3J1.

[2] Faculty of Education, McGill University, 3700 McTavish Street, Montréal, QC. Canada H3A 1Y2.

[3] Montreal General Hospital, 1650 Cedar Avenue, room L9-509, Montréal, QC. Canada H3G 1A4.

[4] Jewish General Hospital Pavilion A-515 3755 Côte-Sainte-Catherine Montréal, QC, H3T 1E2.

  1931-7204/$30.00 https://doi.org/10.1016/j.jsurg.2017.08.024

The IRR between the "Non-Experts" using the rubrics was higher than the IRR of the "Experts" using the traditional surgical approach to assessment. The Cronbach's alpha for the Procedural and Semantic rubrics was 0.94 and 0.97, respectively, indicating very high IRR.
The correlation between the Procedural rubric scores and "Experts" assessment was approximately $r = 0.78$, and that between the Semantic rubric and the "Experts" assessment was roughly $r = 0.66$, indicating high concurrent validity for the Procedural rubric and moderately high validity for the Semantic rubric.

PRINCIPLE CONCLUSION: Clinical reasoning, as measured by some of its "procedural" features, improves over the course of the clerkship year. Rubrics can be created to objectively assess the summary statement of an online interactive trauma VP for "procedural" issues but not for "semantic" issues. Using IRR as a measure, the quality of assessment is improved using the rubrics. The "Procedural" rubric appears to measure changes in clinical reasoning over the course of 3rd-year undergraduate clinical studies. (J Surg Ed 75:779-786. © 2017 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: clinical reasoning, assessment and evaluation, virtual patient, undergraduate medical education, scoring rubric

COMPETENCIES: Medical Knowledge, Practice-Based Learning and Improvement

INTRODUCTION

Clinical reasoning is "a cornerstone of medical practice."[1] The development of clinical reasoning skills permits medical learners to (a) critically analyze patient situations, (b) determine the significance of symptoms, signs, laboratory results and imaging, (c) engage in the development of a hypothesis, or differential diagnosis, (d) participate in active problem solving, and (e) define associated treatment-planning that leads to problem resolution and positive patient outcomes.[2–4] Unfortunately, the direct assessment of clinical reasoning is challenging and particularly difficult during the clerkship year,[5,6] which is especially troubling as many educators anticipate the shift toward competency-based education. At our institution, particularly in the field of trauma, students report a highly variable exposure, few opportunities for independent management of patients, and even fewer opportunities for direct observation and evaluation by supervisors.

In their "Recommendations for Clinical Skills Curricula for Undergraduate Medical Education," the American Association of Medical Colleges suggested that diverse instructional methodologies, including simulated clinical learning opportunities can and should be used to support the development of clinical reasoning skills.[7] VPs are mid-level fidelity simulated online clinical learning opportunities, also described as "multimedia, screen-based interactive patient scenarios"[1] (page 1217) nested in decision-based learning applications that can be used to teach and assess students.[8,9,10,11,12]

Our institution developed a series of trauma VPs based on the ATLS principles of the American College of Surgeons. The VPs include multiple assessment features that can parallel key features examinations, including multiple choice, inquiry assessment (multiple right and wrong answers), and summary statements, simulating "sign-off" to an attending physician. This exercise requires the student to recall the key components of the case, to organize them appropriately and to prioritize treatment procedures creating a clear and contextual analysis of the critical components associated with the VP case. Critical indicators associated with clinical reasoning emphasize learners' ability to synthesize, organize, and prioritize their clinical narrative within a comprehensive, succinct, and clear construct. The summary statement, as an instructional strategy, aligns with these indicators associated with clinical reasoning.[11,13,14]

The summary statement is initiated by an instruction to create the summary as though the student was presenting the case to an attending surgeon. Underneath this instruction is a text input field where the student types their summary. A submit button signifies completion of the summary and the application emails the summary to the instructor. The application then reveals an expert's summary to which the student can compare their own summary.

The Expert Summary

"This is (Name of Student) calling. I am a 3rd-year medical student. I (we) have a young adult male in the trauma bay who as the driver was involved in a T-bone car crash with substantial intrusion where he was sitting. There were no airbags. He smells of alcohol. He was staggering around the accident site when the ambulance arrived, but subsequently became unresponsive. At 3:15, on arrival in the trauma bay, he was somnolent and in severe respiratory distress. He has been intubated. He had a hemopneumothorax and has had a left chest tube inserted. Approximately 600 cc of blood were drained. There are multiple rib fractures and subcutaneous crepitus but he is oxygenating well at this point with a Sat of 90%. He arrived with a pressure of 90 and a pulse of 130 and RR of 30 and shallow. After the chest tube insertion and after approximately 2 L of fluid his pressure is now 100 and his pulse 120. He has been cross-matched for 6 units and we will give him blood as soon as it arrives. He has a closed head injury that has progressed since the accident. His GCS at 4:15 is 5t (E1 V1 M3). Neurosurgery has been called. On exposing and logrolling the patient there is no obvious abdominal injury, however, his left leg is shortened and internally rotated suggesting a left hip fracture. We are about to start the secondary survey and assuming he continues to stabilize we will take him to CT for a total body scan. Is there anything else you would like us to do for the moment?"

The purpose of this study was to develop a tool (the Procedural and Semantic Rubrics) to objectively evaluate the summary statement of a VP and to determine if this approach could measure the evolution of clinical reasoning of medical students during their clerkship year.

METHODS

Study Participants

Two cohorts of 3rd-year McGill medical students were invited to participate in this study. The "Seniors" ($n = 30$) were students from the class of 2016 who were doing their surgery rotation in the final 2 months of their clerkship year (May to June 2015, i.e., end of year). The "Juniors" ($n = 60$) were students from the class of 2017 who were doing their surgery rotation in the first 4 months of clerkship (August to November 2015, i.e., beginning of year). This was the first rotation the Juniors experienced being embedded within a clinical group, taking call and integrating almost entirely within the clinical fora. These circumstances are very different from their initial 2 years of medical school, where a great part of their learning was lecture and laboratory based. The Seniors were also clerkship students similarly embedded within the clinical surgical teams. This was the final rotation before they began their last year of undergraduate medicine. These groups were separated by the experiences of clerkship training and the associated clinical exposure.

All students were required to complete the same online virtual patient case as part of the surgical curriculum. They were advised in writing that their summaries would be stripped of nominative data, reviewed and evaluated anonymously. Nonetheless students needed to consent to have their summaries included in the study. Ethics approval was obtained from the McGill Institutional Review Board.

Virtual Patient and Summary Statement

All students completed the same VP case: a car crash victim with multiple injuries, seen in the hospital's trauma center. The VP was developed by a trauma surgeon (D.F.) and a PhD in education (N.P.) on the DecisionSim™ platform (Kynectiv Inc., Chadsford, PA). The case management required the students' knowledge and ability to apply that knowledge in a clinical setting with the focus being initial management and resuscitation, according to the ATLS principles of the Primary Survey including C-spine control, airway, breathing, circulation, disability, and exposure. In addition, it required the students to recognize and treat respiratory distress due to a pneumothorax. These elements needed to be addressed and completed in the appropriate temporal sequence. The case has extensive branching functionality and therefore multiple outcomes are possible, many of which are determined by the clinical reasoning of the learners. It includes illustrations and animations to reinforce important concepts and is formative in its construct. At the end of the case, each learner was required to summarize the case as though they were reporting to an attending staff person in a real clinical situation: the "summary statement." It consisted of free-text input and was limited to 300 words.

Development of the Procedural and Semantic Rubrics

Two surgeon clinical content "Experts" (D.F. and S.B.) developed an expert summary statement, which served as the gold standard response. It conformed to the Medical Council of Canada's trauma objectives: (a) list and interpret critical clinical findings derived from an appropriate examination performed, in this case, according to the ATLS algorithm completing the primary survey; (b) construct an effective initial management plan to initiate resuscitation of the injured patient and assess the patient's response to resuscitation; (c) list appropriate procedures (e.g., intubation, needle thoracostomy, and chest tube insertion); and (d) interpret investigations useful in the management of the injury.[15] The expert summary statement modeled proper temporal organization, including (a) a list of key concepts underscored in the virtual patient case, and (b) specific semantic qualifiers that would be expected during a sign-off in a real clinical setting.

Based on the expert summary statement, the Procedural and Semantic Rubrics scoring tool was created to evaluate the clinical reasoning of medical students for this VP case. The first rubric, entitled "Procedural" and reproduced in Figure 1, was based on the procedural information necessary to perform a comprehensive ATLS primary trauma survey and the associated management. It was directly related to the AFMC Element Checklist requirement 7.4: "critical judgement/problem solving," a key component of clinical reasoning.[16] Nineteen clinical elements were identified by the surgeons as critical in the management of the VP.

The second rubric, entitled "Semantic" and reproduced in Figure 2, aligned with the Liaison Committee on Medical Education: Standards for Accreditations requirement 7.8: "communication skills, with a specific focus on communication with colleagues."[17] Semantic qualifiers are components of professional communication or terminology that clinicians use, and students are expected to learn to clearly and appropriately communicate a patient's condition in a clinical setting.[13] Twenty-two descriptor words or phrases were included.

Statistical Analysis

The students' summary statements were scored independently by two raters with no training in trauma (N.P. and V.D.), using the Procedural and Semantic rubrics. Each

Organization and Comprehensiveness of Summary Statement		
Indicators:	**Included in Summary Yes/No**	**Score**
1. Student identifies himself or herself		
2. Student identifies the patient		
3. Student identifies the context of the trauma (T-bone car crash) and the cause of injury		
4. Student refers to 'C-spine precautions'		
5. Student refers to Airway		
6. Student refers to subsequent intubation		
7. Student refers to Breathing		
8. Student refers to subsequent needle thoracostomy and chest tube insertion		
9. Student notes the amount of drainage		
10. Student refers to Circulation		
11. Student notes the number of IV's		
12. Student notes the IV solution and the rate of infusion		
13. Student refers to Disability		
14. Student refers to the Glasgow Coma Scale		
15. Student refers to Expose		
16. Student summarizes the severity of the injury at this time		
17. Student includes a time-line		
18. Student refers to the Secondary Survey		
19. The word count is between 150-300		
Score		

FIGURE 1. The Procedural Rubric.

element addressed by the students in the free-text summary statement resulted in the allocation of one point. In the Procedural rubric, if this element was not specifically addressed, but a treatment was identified as necessary, one half of a point was given. For example, if "airway" was not specifically mentioned, but intubation was noted to have occurred, the student received one-half point for addressing airway as part of the ATLS Primary Trauma Survey. Thus raters gave a mark of 0 if the procedure or term was not used, 0.5 mark if it was implied, and one mark if it was clearly mentioned. With respect to the Semantic rubric, specific phrasing was required in the summary statement. If the phrase was included, for example 'severe respiratory distress' or 'pneumothorax' then the learner received one point. If a phrase was missing, a score of zero was given.

a) Inter-rater reliability (IRR)
To determine whether the two rubrics were reliable; that is, whether a student would receive the same score independent of who did the rating, Cronbach's alpha, a measure of IRR, was computed for each scale separately using the two "Non-Expert" raters' scores. Because IRR was high for both rubrics, in further analyses, the two raters' scores for each rubric were averaged; resulting in one Procedural and one Semantic score for each student.

b) Concurrent validity of the rubrics
Concurrent validity (a numerical indication of how well a new scale measures the same underlying factor compared to a trusted scale),[18] was computed for both rubrics by comparing the combined raters' scores to the established traditional method of evaluation. This was a subjective assessment, similar to what a clinician would do when asked if a student's sign-off was acceptable. In order to accomplish this, the summary statements from both cohorts were independently assessed by the two clinical "Experts." This "Expert" subjective assessment was based on a 'quick read,' a traditional surgical method of assessment. Both "Experts" (D.F. and S.B.) independently read each summary statement and rated it as acceptable (pass), not acceptable (fail) or partially acceptable (borderline). The IRR for the two "Experts" assessments was high, with a Cronbach's alpha of 0.846. The "Experts'" judgements were therefore combined, such that a summary was classified as "Pass" if both "Experts" judged the summary to be a pass, a "Fail" if both "Experts" marked it as a fail, and as "Borderline" if either or both "Experts" marked the summary as "borderline" or if there was disagreement between the two "Experts." Because the "Expert" assessment (pass, borderline or fail) was an ordinal scale, a Kendall τ was

Use of Semantic Quantifiers or Structures		
Semantics – words that should be included:	**Included in Summary Yes/No**	**Score**
1. 'T-bone'		
2. 'Substantial intrusion'		
3. 'Severe respiratory distress'		
4. 'Hemopneumothorax' or pneumothorax'		
5. Multiple rib fractures		
6. 'Subcutaneous crepitus' or 'Bony crepitus'		
7. 'Oxygenating' and 0_2 SATs		
8. 'Somnolent'		
9. Inclusion of Pressure and VS x 1		
10. Inclusion of Pressure and VS x 2		
11. Inclusion of time of arrival		
12. Inclusion of time when GCS given		
13. 'Cross match with number of units'		
14. 'Hemodynamically unstable'		
15. 'Closed head injury that has progressed'		
16. 'Unresponsive with time'		
17. Reports Glasgow Coma Scale as '5t' and/or E1 V1 M3		
18. 'Internal rotation and shortening of left leg'		
19. States that 'Neurosurgery has been called' refer to specialized care		
20. States that 'Secondary survey will now be started'		
21. States that patient will be sent CT for total body scan		
22. Requests feedback from attending staff		
Score		

FIGURE 2. The Semantic Rubric.

computed to compare the distribution of "Experts" subjective ratings between the two cohorts (Junior or Senior). Each rubric was subjected to a separate one-way ANOVA with the "Experts" subjective rating (Pass, Borderline, or Fail) as the grouping variable and the rubric score as the outcome. Between-group differences were tested using Fisher's Least Significant Difference (L.S.D.) test.

c) Construct validity

To determine whether the Senior and Junior scores differed significantly, independent group *t*-tests were conducted separately for both the Procedural and Semantic rubrics. Cohen's *d* was used as a measure of effect size.

RESULTS

Twenty-five of 30 eligible students participated in the Senior group and 57/60 students participated in the Junior group.

a) IRR

The Cronbach's alpha for the Procedural and Semantic rubrics was 0.94 and 0.97, respectively, indicating very high IRR.

b) Concurrent validity

According to the "Experts'" subjective assessments, in the Senior group, 8/25 (32%) of students passed, 5/25 (20%) students failed, and 12/25 (48%) of students were given a borderline. In the junior group, 12/57 (21%) students passed, 23/57 (40%) students failed, and 22/57 (39%) students were given a borderline. There was no statistical difference in the marks given to the Senior versus Junior groups by the "Experts" (Kendall's $\tau = 0.182$, $p = 0.07$).

The Procedural rubric ANOVA was statistically significant ($F_{(2,79)} = 61.73$, $p < .001$, eta-squared $= 0.61$, with all group differences significant by Fisher's L.S.D. at $p < .001$). The Semantic rubric was also statistically significant ($F_{(2,79)} = 35.00$, $p < .001$, eta-squared $= 0.47$, with all group differences significant by Fisher's

L.S.D. at $p < 0.001$). For determination of concurrent validity, the eta-squared statistic can be read as an approximation of r^2, the percent of variance explained, indicating that the correlation between the Procedural rubric scores and "Experts" assessment was approximately $r = 0.78$, and that between the Semantic rubric and the "Experts" assessment was roughly $r = 0.66$, indicating high concurrent validity for the Procedural rubric and moderately high validity for the Semantic rubric.

(c) Construct validity

The average Procedural rubric score for the Senior group was 59% ± 13% while for the junior group, it was 51% ± 12% ($t(80) = 2.715$; $p = 0.008$; Cohen's $d = 1.53$. The average Semantic rubric score for the Senior group was 31% ± 15% while for the Junior group, it was 28% ± 14% ($t(80) = 1.010$; $p = 0.316$, ns).

DISCUSSION

While the evolution of clinical reasoning has been studied,[9,4,10,19] it is difficult to assess clinical reasoning during clerkship: This is a result of a variety of issues, which include, but are not confined to (a) logistics: students are often situated in diverse clinical settings, (b) availability of learning opportunities: all students are not situated in areas that provide direct and critical on-site clinical experiences, and (c) assessment practices: global ratings for end-of-rotation medical students are infrequent, inadequate and removed from those who have directly observed individual student's behaviors.[6]

In this context, virtual patients are considered acceptable surrogates to real clinical encounters by regulatory bodies including the Medical Council of Canada,[15] and the American Association of Medical Colleges Liaison Committee for Medical Education.[17] VPs allow learners to interactively "work through the steps of diagnosing and managing patients and making clinical decisions without the risk of causing patient harm"[20] (page 1713). They are problem-oriented, self-directed, and effective in teaching clinical domain knowledge and clinical reasoning[21] and can support faculty assessment of a student's progress through the provision of formative and summative assessment. They can also fulfill the requirement of providing equal learning opportunities across the various teaching sites.

This study was conducted to determine whether rubrics could be created to objectively assess the summary statement of an online interactive VP and to measure a change in certain aspects of clinical reasoning over time. Two rubrics were created to assess the "summary statement" created by the students at the completion of a VP case. The first, the Procedural rubric, specifically looked at domain knowledge as a component of clinical reasoning, addressing the ability to organize, detail, and comprehensively present a patient case. The second, the Semantic rubric, looked at the use of professional communication that would be anticipated during this case presentation.

Berman et al.[21] highlighted the potential for the use of applied analytics to reflect a student's understanding of key findings, as well as the assessment and identification by faculty of mastery learning. These summary statements can be used to objectively assess (a) the ability to organize and comprehensively describe the temporal sequence of events associated with a patient,[13] (b) whether key learning objectives as highlighted in the case introduction align with critical case summary statements expected at the end of the case, and (c) the utilization of appropriate abstract semantic qualifiers. These are described as verbal or written representations of professional terminology used by clinicians to clearly and precisely communicate a clinical situation[14,22] and "can be assessed as a marker for clinical reasoning"[5] (page 94).

In this study, both rubrics used to evaluate the summary statements were found to have very high IRR, suggesting that this tool is robust enough to allow different raters, even those without any knowledge of trauma, to evaluate summary statements with very comparable results. The tool was compared to how clinicians would usually evaluate a clinical case presentation, the historical standard being a read of the summary (or in a real clinical situation listening to the summary) and deciding if it was acceptable, borderline, or unacceptable. This subjective assessment was added to a global impression of success or failure for the rotation. There was very good correlation between the two "Expert" subjective assessments and the two rubrics, signifying concurrent validity and suggesting that this form of student evaluation is viable and could be used to replace some of the existing less objective components of the students' global assessment.

When comparing the two student groups, results from the Procedural rubric determined that students at the end of their 3rd year scored higher than those at the beginning of the year. Results from the Semantic rubric did not demonstrate a difference. The failure to find significant student progress with the Semantic rubric may suggest that only the Procedural rubric has construct validity. On the other hand, it may also indicate that in the 3rd year of clerkship, students do not acquire the correct terms and phrases or the ability to use them properly in a medical dialog. It seems possible that greater emphasis placed on teaching the procedures evaluated by the Procedural rubric, is being done at the expense of the "softer," but equally important, medical semantics and communication skills within the trauma clinical setting and are left to be honed simply through practice and conversation. The use of professional semantics is one of a variety of methods to support clarity, and to prevent miscommunications or misunderstandings. This has considerable implications,

particularly as students become residents who are moving toward shift-based working models with more frequent verbal "sign-offs" and more opportunities to make errors in communication.[23–25] This may also signal a need for faculty to include semantics as part of their direct instruction.

The fact that, unlike the Procedural rubric, the "Expert" assessment could not distinguish between the Junior and Senior groups, suggests that the Procedural rubric may be more sensitive to student progress. The objective assessment of student learning can provide information that is not easily obtained from traditional subjective "Expert" assessment and to combine the use of rubrics and VPs affords new opportunities in this regard.

Although the Seniors scored better than the Juniors, their scores were still relatively low. This is likely due to the fact that the rubrics were created by experts with many years of trauma experience and even the Senior students would not be expected to approach this level of performance. It would be interesting to extend this exercise to residents expected to be at higher levels of performance. It is also noteworthy that the Experts' subjective assessments (pass, borderline, and fail) had concordance with the rubrics. This might be explained by the heuristics developed by the surgeons over the years, making the apparently "subjective" somewhat more "objective" than it might seem to be on the surface.

The role of the rubric-based objective assessments cannot, and is not meant to replace clinical evaluations, however it provides additional information that contribute to the global evaluations. As well, the rubrics could potentially be used by non-clinical administrative personnel freeing the clinicians from the assessment process and allowing them more teaching time.

VPs are still a relatively new learning methodology and research on the implications of the use, as well as on the value to the medical school's assessment capabilities is still unfolding. The use of objective rubrics may permit clinical supervisors to better identify marginal students and specify their areas of weakness. These issues are of substantial importance at our institution, especially in undergraduate trauma education. With high-volume multi-system trauma only available at 1 of 5 sites, trauma exposure for the students is highly variable. To standardize exposure, fulfill accreditation requirements and improve assessment of clinical reasoning in trauma, the use of virtual patients and of their summary statement with properly validated evaluation tools will be paramount.

There are some limitations to this study that should be acknowledged. This study was based on a single case that focused on trauma. The relatively short academic time between cohorts may not have been sufficient to allow measurable progress for semantics. The cohorts were both comprised of 3rd year undergraduate medical students. For these reasons, we feel that the results are not generalizable. Further studies should address different VP cases and other domains, both within surgery and beyond, longer time intervals between measurements should be considered and a broader mix of student's groups may give more insight into the evolution of clinical reasoning skills. Questions about the applicability of this methodology to assess post-graduate students could be addressed in future studies.

CONCLUSION

Clinical reasoning, as measured by some of its "procedural" features, improves over the course of the clerkship year. Rubrics can be created to objectively assess the summary statement of an online interactive trauma VP for "Procedural" issues but not for "Semantic" issues. Using IRR as a measure, the quality of assessment is improved using the rubrics. The "Procedural" rubric appears to measure changes in clinical reasoning over the course of 3rd-year undergraduate clinical studies.

ACKNOWLEDGMENT

We would like to thank Kynectiv Inc. for the use of their platform DecisionSim™ to house the virtual patient and to provide student access as an "in-kind" contribution to this project.

We would also like to thank Valerie Dory at the McGill University Center for Medical Education for her help in reviewing the summaries and for doing the initial statistical analysis.

REFERENCES

1. Berman NB, Durning SJ, Fischer MR, Huwendiek S, Triola M. The role for virtual patients in the future of medical education. *Acad Med*. 2016;91:1217-1222.
2. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2004;39:98-106.
3. Fonteyn ME, Ritter BJ. Clinical reasoning in nursing. Profession. Higgs J, Jones M, editors. *Clinical Reasoning in the Health*. 3rd ed. Oxford, UK: Oxford, Butterworth Heinemann, 2008.
4. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;84:1022-1028.
5. 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*. 2016;91:94-100.
6. Kogan JR, Holmboe ES, Hauer KE. Tool for direct observation and assessment of the clinical skills

of medical trainees. *J Am Med Assoc*. 2009;302:1316-1326.

7. Association of American Medical Colleges. Recommendations for Clinical Skills Curricula for Undergraduate Medical Education Available at: https://www.aamc.org/download/130608/data/clinicalskills_oct09.qxd.pdf.pdf. Accessed February 2017.

8. Ellaway R, Candler C, Greene P, Smothers V. *An architectural model for Medbiquitous virtual patients*. 2006. Available at: http://groups.medbiq.org/medbiq/display/VPWG/MedBiquitous+Virtual+Patient+Architecture Accessed March 2017.

9. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ*. 2005;39:418-427.

10. Andersen D. How can educators use simulation applications to teach and assess surgical judgment? *Acad Med*. 2012;87:1-8.

11. Bateman J, Allen ME, Kidd J, Parsons N, Davies D. Virtual patient design and its effect on clinical reasoning and student experience: a protocol for a randomized factorial multi-centre study. *BMC Med*. 2012;12:62.

12. Bateman J, Allen M, Samani D, Kidd J, Davis D. Virtual patient design: exploring what works and why. A grounded theory study. *Med Educ*. 2013;47:595-606.

13. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *NEJM*. 2006;355:2217-2225.

14. Posel N, McGee JB, Fleiszer D. Twelve tips to support the development of clinical reasoning skills using virtual patient cases. *Med Teach*. 2015;37:813-818.

15. Medical Council of Canada Objectives for the Qualifying Examination. 3rd Edition 3.3.16. Available at: http://apps.mcc.ca/Objectives_Online/objectives.pl?lang=english&loc=contents#expert Accessed March 2017.

16. Association of Faculties of Medicine of Canada: Elements Checklists: Standard 8—Curricular Management, Evaluation and Enhancement Available at: https://afmc.ca/accreditation/interim-accreditation-review-process-irp/checklists. Accessed February 2017.

17. Liaison Committee on Medical Education. Function and Structure of a Medical School: Standards for Accreditation of Medical Education Programs Leading to the M. D. Degree. Available at: https://med.virginia.edu/ume-curriculum/wp-content/uploads/sites/216/2016/07/2017-18_Functions-and-Structure_2016-03-24.pdf. Accessed February 2017.

18. McIntire SA, Miller LA. *Foundations of Psychological Testing*. 2nd Edition Sage Publishing Co., 2005.

19. Dror I. How to exploit cognitive opportunities in medical education: Improving patient care and safety. Flanders Family Visiting Professor in Medical Simulation. *McGill University*. 2012 September 13, 2012.

20. Lang VJ, Kogan J, Berman N, Torre D. The evolving role of online virtual patients in internal medicine clerkship education nationally. *Acad Med*. 2013;88:1713-1718.

21. Berman N, Fall LH, Smith S, Levine DA, Maloney CG, Potts M, Siegel B, Foster-Johnson L. Integration strategies for using virtual patients in clinical clerkships. *Acad Med*. 2009;84:942-949.

22. Bordage G, Lemieux M. Semantic structures and diagnostic thinking of "Experts" and novices. *Acad Med*. 1991;66:S70-S72.

23. Horowitz L, Moin T, Green M. SIGNOUT? Development and implementation of an oral sign-out skills curriculum *J Gen Intern Med*. 2007;22:1470-1474.

24. Wolpaw TM, Wolpaw DR, Papp KK. SNAPPS: a learner-centered model for outpatient education. *Acad Med*. 2003;78:893-898.

25. Leonard M, Bonacum D, Graham S. *SBAR technique for communication: a situational briefing model*. 2013. Available at: http://www.ihi.org/knowledge/Pages/Tools/SBARTechnqiueforCommunicationASituationalBriefingModel.aspx. Accessed May 9 2013.