



Improving Passage Rate on USMLE Step 2 Clinical Skills: Results from a Pilot Program

Michael S. Ryan¹ · Melissa Bradner² · Fidelma Rigby³ · Bennett Lee⁴ · Elizabeth Waterhouse⁵ · Catherine Grossman⁴

Published online: 15 July 2019

© International Association of Medical Science Educators 2019

Abstract

In the United States (US), successful passage of United States Medical Licensing Examination (USMLE) Step 2 Clinical Skills (Step 2 CS) is required to enter into residency training. In 2017, the USMLE announced an increase in performance standards for Step 2 CS. As a consequence, it is anticipated that the passage rate for the examination will decrease significantly for both US and international students. While many US institutions offer a cumulative clinical skills examination, their effect on Step 2 CS passage rates has not been studied. The authors developed a six-case, standardized patient (SP)-based examination to mirror Step 2 CS and measured impact on subsequent Step 2 CS passage rates. Students were provided structured quantitative and qualitative feedback and were given a final designation of “pass” or “fail” for the practice examination. A total of 173 out of 184 (94.5%) students participated in the examination. Twenty SPs and \$26,000 in direct costs were required. The local failure rate for Step 2 CS declined from 4.5% in the year preceding the intervention to 2.1% following the intervention. In the same timeframe, the US failure rate for Step 2 CS increased from 3.8 to 5.1%, though the difference between local and national groups was not significantly different ($P = .07$). Based on the initial success of the intervention, educational leaders may consider developing a similar innovation to optimize passage rates at their institutions.

Keywords Step 2 CS · Clinical skills · Standardized patients · Assessment

Introduction

The United States Medical Licensing Examination (USMLE) Step 2 Clinical Skills (Step 2 CS) was designed to assess medical students’ abilities to apply medical knowledge, communication, and examination skills to the care of patients [1].

Following its introduction in 2004, several studies examined validity evidence for the examination [2–4]. In addition to finding validity for the content of the examination, initial and subsequent studies demonstrated correlation between scores on Step 2 CS and performance in residency [4, 5]. As a result, successful passage of Step 2 CS became a requirement for entry into residency training and medical licensure in the United States (US).

Since 2005, the Step 2 CS passage rate has ranged 95–98% for US medical school graduates and 70–84% for non-US medical school graduates. In recent years, the passage rate for non-US graduates has remained relatively stable between 78 and 81% [6]. In 2017, the USMLE announced an increase in performance standards for the examination beginning in September of that year. The USMLE projected that, if applied to recent test takers, these performance standards would result in a substantial decrease in Step 2 CS passage rates for both US and international medical school graduates [7].

Prior to the introduction of Step 2 CS, many US medical schools already offered a local cumulative assessment of students’ clinical skills following the core clerkships [8]. After the introduction of the Step 2 CS requirement, additional schools

✉ Michael S. Ryan
michael.ryan1@vcuhealth.org

¹ Department of Pediatrics, Virginia Commonwealth University School of Medicine, 1201 E. Marshall St, Suite 4-200, Box 980565, Richmond, VA 23298-0565, USA

² Department of Family Medicine and Population Health, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

³ Department of Obstetrics and Gynecology, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

⁴ Department of Internal Medicine, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

⁵ Department of Neurology, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

introduced comprehensive assessments while others either continued their existing assessments or made some modifications to better prepare students for the national examination [9, 10].

Two author groups have described the relationship between components of in-house examinations and performance on similar components in Step 2 CS. In both cases, a modest positive correlation was observed between components of the in-house cumulative examinations and Step 2 CS [11, 12]. While correlations are valuable in supporting the validity of in-house examinations, students may be more interested in whether the examination ultimately promotes passage of Step 2 CS. Only one previous study has reported on the passage rates associated with a formal Step 2 CS preparation course. That course demonstrated improved confidence and perceived competence in preparation for Step 2 CS among students from an international medical school. The passage rates increased in that cohort but also increased in all international medical school graduates during that same time period [13].

To address local concerns over Step 2 CS passage rates, we developed a cumulative clinical skill assessment conducted at the conclusion of the core clerkships. The purpose of this standardized patient (SP)-based assessment was to specifically prepare students for the Step 2 CS examination and as a consequence, improve local passage rates. In this manuscript, we provide a description of the clinical skills assessment and the initial outcomes on Step 2 CS passage rates following its implementation.

Materials and Methods

Setting and Participants

The study took place at the Virginia Commonwealth University School of Medicine (VCU-SOM), a public, urban, academic medical center in Richmond, Virginia, in the US. In the study period, VCU-SOM enrolled approximately 185 students to the Doctor of Medicine program on the main campus for each academic year. Prior to the intervention, all students were required to pass a cumulative clinical skills examination prior to entry into the core clerkship phase. Three clerkships also required passage of observed structured clinical examinations (OSCE) to pass the respective clerkship. However, we did not provide a cumulative clinical skills examination at the conclusion of the clerkship phase. For the 2016–2017 academic year (Class of 2018), we developed and piloted a six-station standardized patient (SP)-based OSCE (Practice Step 2 CS) conducted at the VCU-SOM.

Case Development

We incorporated principles of experiential learning [14] to guide the organization of the examination. We designed our examination to mirror Step 2 CS in terms of content, logistics,

and evaluation methods to provide our learners with as authentic an experience as possible. Overall goals and a content blueprint were developed based on a review of publically available material provided by the USMLE and a student-oriented study guide [1, 15]. Individual cases were then developed and/or adapted from three primary sources: MedEdPORTAL [16], internally-developed OSCEs, and the student-oriented study textbook [15]. A total of 6 cases was created to highlight the major clinical disciplines and to represent exactly half the number of cases assessed during Step 2 CS. Each case was first drafted by two authors (MSR and CG) and then assigned to a clerkship director(s) for review. The final cases each included key learning objectives, counseling tasks, and critical elements of history, physical, assessment, and workup. These details are summarized in Table 1.

Scoring

Scoring rubrics were developed to assess each of the Step 2 CS subcomponents beside English language proficiency. English language proficiency was not explicitly assessed because it was a requirement for matriculation at VCU-SOM and because we had not observed a failure for this Step 2 CS component. To assess communication skills, we used a previously described framework for teaching and assessing this skill [17]. Data gathering was assessed using a case-specific checklist completed by the SP. The encounter note was graded using a case-specific rubric developed by the authors of the case and revised by the respective clerkship director. A random sample of 10 notes was selected to iteratively develop the encounter note rubric. Once developed, encounter notes were assigned to each clerkship director(s) such that the specialty of the director corresponded with primary discipline emphasized in the case.

Administration

The Practice Step 2 CS was run using similar resources as would be used for OSCEs at our institution. We ran all six stations in a 3-hour time frame using a structure mirrored after Step 2 CS: student were given 15 minutes of encounter time, 10 minutes designated note writing time, and then 5 minutes to allow migration to the subsequent encounter room. Cell phones and other electronic devices were prohibited during the exam and hallway proctors were used to enforce an exam-like environment. Computer stations outside each encounter room were equipped for computer entry for patient note forms for each encounter. The forms were structured to resemble the published encounter note form template provided on the USMLE website.

Our physical infrastructure allowed us to run 2 identical 6 station OSCEs at a time with 12 students. We hired and trained

Table 1 Summary of cases developed for Practice Step 2 CS

Case and primary discipline ^a	Objectives	Counseling tasks	Critical elements of history and physical	Critical elements of assessment	Critical elements of workup
Chest pain, Internal Medicine	Formulate a differential diagnosis for a middle-aged adult presenting with chest pain Differentiate between cardiac and non-cardiac chest pain	Respond to a patient asking, “am I having a heart attack?” Provide smoking cessation counseling	Character of chest pain Coronary artery disease risk factors Complete cardiovascular examination	Chest pain (cardiac vs. non-cardiac)	Rule out myocardial infarction
Back pain, Neurology	Formulate a differential diagnosis for an adult with low back pain Negotiate pain medication options	Respond to a patient requesting opiates Provide alcohol counseling	CAGE questionnaire Musculoskeletal examination Neurological examination	Acute back pain (neurological vs. musculoskeletal)	Avoid high-cost/high radiation testing as first line
Epigastric Pain, Surgery	Formulate a differential diagnosis for an adult male presenting with acute abdominal pain Develop an initial workup plan for an acute abdomen	Respond to a patient asking “do I need surgery?” Provide alcohol counseling	Acuity of onset Peritoneal signs CAGE questionnaire	Acute abdomen	Acute series “Abdominal labs”
Fatigue, Psychiatry	Formulate a differential diagnosis for an adult presenting with fatigue Consider primary psychosocial etiologies for somatic complaints	Counsel a patient with depressive symptoms Screen for substance abuse	Depression screen (i.e., SIG E CAPS) Social history	Depression vs. Substance Abuse	n/a
Menorrhagia, Obstetrics and Gynecology	Formulate a differential diagnosis for a pre-menopausal woman presenting with menorrhagia Create an initial workup plan for menorrhagia	Counsel regarding safe sex practices	Women’s health history Sexual history Abdominal exam	Causes of menorrhagia in pre-menopausal woman	Request pelvic examination
Jaundice, Pediatrics	Formulate a differential diagnosis for an infant presenting with jaundice Triage a phone call appropriately	Respond to a parent asking, “do I need to take my child to the ER?”	Birth history Diet and growth	Differential diagnosis for low-risk infant with jaundice	Routine follow-up only

^a A specific case for Family Medicine was not developed because the authors felt each case may be expected in a Family Medicine office

20 SPs for this event to fill the required schedule which spanned a 2-week time frame.

All students were scheduled for a half-day to participate in Practice Step 2 CS at the conclusion of their clerkship year (late February through early March). Scheduling was coordinated in conjunction with the respective clerkship teams to minimize time away from clinical activities.

Score Reporting to Students

Scores from all components were compiled and reported back to individual students. The format included the percentage of points received for each component, the comparison of that student to the mean, a summary of comments provided by the SPs, and an assignment of “pass” or “fail” based on a cumulative score of 70% for each component (communication and integrated clinical encounter). If a student received a score of $\leq 70\%$ for either component, they received a final designation of “fail” for the examination. The cutoff of 70% was chosen because it represented the standard cutoff for pass/fail throughout our curriculum.

All students were provided written feedback from one author (MSR) prior to taking the Step 2 CS examination. As this was formative, remediation was not required, but students were offered counseling to discuss methods to improve performance on Step 2 CS. Outside of requiring a passing Step 2 CS examination prior to graduation, we did not require that students complete the examination at any specific time during medical school. Most students took the examination in July (20.9%) or August (35.4%) of their final year while a minority took the examination prior to May (10.2%) or after September (12.6%).

Results

Outcomes were measured through review of resources required, performance on the Practice Step 2 CS examination, student reactions, and performance on Step 2 CS. This protocol was approved exempt by the VCU Institutional Review Board.

Resources

A total of 173 out of 184 eligible (94.5%) students participated in the Practice Step 2 CS in the first iteration. Event training required approximately 45 cumulative SP hours (no. of individual SP hours \times training hours). The total SP event time (no. of individual SP hours \times event hours) required during the event was approximately 576 h. Approximate total SP costs for the event were \$26,000. The total approximated time required for physician grading of all encounter notes was estimated at 174 hours.

Examination Performance

More than half of students “failed” the Practice Step 2 CS due to scoring below 70% for communication, integrated clinical encounter data, or both. Sixty-one students (35.2%) passed both components, seventy-eight (45%) passed one component but failed the other, and thirty-four (19.7%) failed both components. Four students failed Step 2 CS; two of whom failed one component of the Practice Step 2 CS and two of whom failed both components of Step 2 CS. All students passed Step 2 CS if they passed both components of the Practice Step 2 CS.

Student Reactions

We asked students to reflect on the value of the Practice Step 2 CS after they completed the Step 2 CS using a combination of quantitative and qualitative measures. First, learners were asked to rate the overall value of the Practice Step CS by responding to the prompt, “do you feel like the mock CS (Practice Step 2 CS) prepared you for the real exam (1–5, 1 = not at all to 5 = very much)?” Because the examination was initiated as a pilot, we next asked learners to indicate suggestions for modifying the content and/or feedback processes for future iterations. For these questions, learners were provided with a series of response options and an opportunity to select “other” to provide an open-ended response. We then asked learners to suggest whether the examination should be offered again the following year. Finally, we provided an opportunity for any additional open-ended comments.

Responses were provided by 63 individuals. Overall, students reported the Practice Step 2 CS prepared them well for Step 2 CS. A majority of students (59.5%) provided scores of “4” or “5” when rating whether the Practice Step 2 CS prepared them for the examination. Nearly all (98.4%) of students felt the Practice Step 2 CS should be offered to subsequent classes. Seventy-one percent of students felt the most valuable component of the examination was “just having the opportunity to participate in a mock CS exam.” Components such as “receiving a ‘pass/fail’ score,” a “written summary of my performance,” and “comments from the SPs” were deemed valuable but less critical by respondents. When asked for constructive feedback on what, if anything, could be deleted in subsequent iterations, 36% reported “providing a ‘pass/fail’ designation.” However, 34% reported “keep it all.” Open-ended comments revealed that students felt the Practice Step 2 CS was more difficult than Step 2 CS, that the feedback took too long, and that failure induced anxiety for the real examination.

Effect on Performance

We measured the effect of the Practice Step 2 CS by comparing student performance in the previous two academic years

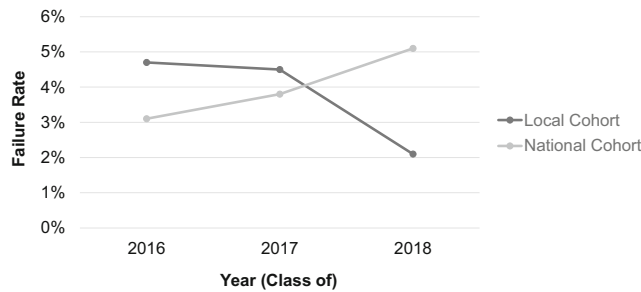


Fig. 1 Step 2 CS performance trends for Class of 2016–2018

with that of our intervention group. We also compared the rates of pass vs. fail in consecutive cohorts of our students as well as the national data using chi-square to make comparisons between categorical variables. A small number of students did not participate in the OSCE due to conflicting vacation time or competing clinical responsibilities.

However, we analyzed our data using intention-to-treat; all students who were eligible for the Practice Step 2 CS were included in intervention group regardless of whether they took the practice examination.

Overall, the absolute failure rate of Step 2 CS decreased for students who received our intervention. The failure rate declined from 4.7% (Class of 2016) and 4.5% (Class of 2017) to 2.1% (Class of 2018). By comparison, the national failure rate *increased* over that same interval from a low of 3.1% (Class of 2016) to 5.1% (Class of 2018). Figure 1 illustrates this trend.

The difference between performance for our local cohort and the national cohort was not significantly different for the Class of 2016 ($\chi^2 = 1.395, P = .24$) or Class of 2017 ($\chi^2 = 0.255, P = .61$). For the Class of 2018 cohort, there was a trend toward a significant difference in comparing our cohort with the national cohort ($\chi^2 = 3.42, P = .07$). A summary of the pass/fail rates and chi-square analysis is provided in Table 2.

By comparison, mean scores remained relatively unchanged and comparable to the national trends on USMLE Step 2 Clinical Knowledge in the same time frame. During the study period, mean Step 2 Clinical Knowledge scores for VCU-SOM ranged from 243 to 246 (STD 16–17) while mean performance among the national comparison groups were 242–243 (STD 17).

Discussion

We developed a six-station SP-based OSCE to improve passage rate on USMLE Step 2 CS at our institution. Despite increased performance standards for the Step 2 CS [6] we observed a decline in our failure rate, when compared to both local and national cohorts following the implementation of the Practice Step 2 CS. Based on student responses, the opportunity to practice an examination that mirrored the real exam in terms of timing, requirements, and grading provided the primary benefit to our students. Students felt as though the intervention prepared them well for the examination; however, they offered several suggestions to improve the experience for future classes.

Following student feedback, we made several changes to the Practice Step 2 CS for the subsequent year. First, we incorporated a large group review of the encounter notes in lieu of requiring clerkship directors to provide grades for each note. The individual grading, while valuable, was time intensive and the rate limiting factor in providing students with feedback in an appropriate timeframe. For the Class of 2019 cohort, we offered 4 large group sessions in which one or both case developers reviewed each of the objectives, key findings, and rubrics for each note. Students were provided electronic copies of their notes during this session to grade on their own. Next, we incorporated increased emphasis on counseling in both our OSCE cases and with debrief sessions to students. This was based on feedback students provided after completing the Step 2 CS examination. We also made several modifications to the SP instructions and grading rubric for the epigastric pain/surgical case acknowledging that performance on that case appeared to be an outlier. Finally, we have continued to provide written feedback summarizing communication scores provided by the SPs. However, we did not continue to provide a pass/fail designation to students as that was deemed less helpful from the student respondents.

We plan on continuing to offer the Practice Step 2 CS for our rising fourth year students in subsequent classes. We will analyze performance on Step 2 CS for our current cohort to determine if the time-saving changes for faculty resulted in any detrimental impact for our students.

Table 2 Passage and Failure Rates on Step 2 CS for Medical Students from 2016 to 2018

Year (Class of)	Local cohort			National cohort ^a			Statistics	
	Pass (n)	Fail (n)	Fail %	Pass (n)	Fail (n)	Fail %	Chi-square	P value
2016	164	8	4.7%	18761	597	3.1%	1.40	0.24
2017	169	8	4.5%	18979	748	3.8%	0.26	0.61
2018	183	4	2.1%	18932	1023	5.1%	3.42	0.07

^a This data differs slightly from publically available performance data for Step 2 CS. The NBME provided data to the authors from the national cohort corresponding to the same timeframe as the VCU-SOM academic year to allow head to head comparisons

Ultimately, we aim to incorporate performance on the Practice Step 2 CS to decisions regarding promotion and advancement. However, doing so would result in a dramatic shift in the stakes of the examination. While our cases and rubrics were developed to optimize content validity through use of blueprints and availability literature; further validity evidence would be required if we shifted from a low-stake formative assessment to a relatively high-stake summative assessment [18]. This would necessitate a thorough review of the cases and the grading rubrics to ensure inter-rater reliability in standardized patients and encounter note raters (internal structure validity) and familiarity with the structure of the documentation system (response process validity). In addition, we would need to create additional cases to ensure the integrity of the assessment and prevent feed forward practices among students.

Conclusions

Overall, the results of this pilot suggest a dedicated Practice Step 2 CS may counteract the increasing failure rates projected in the near future for both US and non-US medical schools. Based on the initial success of the intervention, educational leaders may consider developing a similar innovation to optimize passage rates at their institutions.

Acknowledgments The authors would like to thank the clerkship directors and simulation center staff at VCU who assisted in the construction of cases, administration of the simulation, and grading of encounter notes described in this manuscript.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This study was deemed exempt by the Virginia Commonwealth University IRB. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent A statement regarding “informed consent” is not applicable.

References

1. United States Medical Licensing Examination. Step 2 Clinical Skills: content description and general information. 2017. <http://www.usmle.org/pdfs/step-2-cs/cs-info-manual.pdf>. Accessed 17 Dec 2017.
2. De Champlain A, Swygert K, Swanson DB, Boulet JR. Assessing the underlying structure of the United States Medical Licensing Examination Step 2 test of clinical skills using confirmatory factor analysis. *Acad Med*. 2006;81(10 Suppl):S17–20.
3. Harik P, Clauser BE, Grabovsky I, Margolis MJ, Dillon GF, Boulet JR. Relationships among subcomponents of the USMLE Step 2 Clinical Skills Examination, the Step 1, and the Step 2 Clinical Knowledge Examinations. *Acad Med*. 2006;81(10 Suppl):S21–4.
4. Taylor ML, Blue AV, Mainous AG, Geesey ME, Basco WT. The relationship between the National Board of Medical Examiners' prototype of the Step 2 clinical skills exam and interns' performance. *Acad Med*. 2005;80(5):496–501.
5. Cuddy MM, Winward ML, Johnston MM, Lipner RS, Clauser BE. Evaluating validity evidence for USMLE Step 2 Clinical Skills Data Gathering and Data Interpretation Scores: does performance predict history-taking and physical examination ratings for first-year internal medicine residents? *Acad Med*. 2016;91(1):133–9.
6. National Board of Medical Examiners. Performance Data. 2019. <https://www.usmle.org/performance-data/>. Accessed 4 Jan 2019.
7. United States Medical Licensing Examination. Change in performance standards for Step 2 CS. 2017. <http://www.usmle.org/announcements/default.aspx?ContentId=210>. Accessed 13 Dec 2017.
8. Hauer KE, Hodgson CS, Kerr KM, Teherani A, Irby DM. A national study of medical student clinical skills assessment. *Acad Med*. 2005;80(10 Suppl):S25–9.
9. Gilliland WR, La Rochelle J, Hawkins R, et al. Changes in clinical skills education resulting from the introduction of the USMLE step 2 clinical skills (CS) examination. *Med Teach*. 2008;30(3):325–7.
10. Hauer KE, Teherani A, Kerr KM, O'Sullivan PS, Irby DM. Impact of the United States Medical Licensing Examination Step 2 Clinical Skills exam on medical school clinical skills assessment. *Acad Med*. 2006;81(10 Suppl):S13–6.
11. Berg K, Winward M, Clauser BE, Veloski JA, Berg D, Dillon GF, et al. The relationship between performance on a medical school's clinical skills assessment and USMLE Step 2 CS. *Acad Med*. 2008;83(10 Suppl):S37–40.
12. Dong T, Swygert KA, Duming SJ, Saguil A, Gilliland WR, Cruess D, et al. Validity evidence for medical school OSCEs: associations with USMLE® step assessments. *Teach Learn Med*. 2014;26(4):379–86.
13. Levine RB, Levy AP, Lubin R, Halevi S, Rios R, Cayea D. Evaluation of a course to prepare international students for the United States Medical Licensing Examination step 2 clinical skills exam. *J Educ Eval Health Prof*. 2017;14:25.
14. Kolb D. *Experiential learning: experience as the source of learning and development*. Englewood Cliffs: Prentice-Hall; 1984.
15. Le T, Bhushan V, Sheikh-Ali M, Lee K. *First aid for the USMLE Step 2 CS (clinical skills)*. 4th ed. New York: McGraw-Hill Medical Publishing Division; 2012.
16. Akins R. Victor Billar, acute abdominal pain. *MedEdPORTAL*. 2012;8:9118. https://doi.org/10.15766/mep_2374-8265.9118.
17. Makoul G. The SEGUE framework for teaching and assessing communication skills. *Patient Educ Couns*. 2001;45(1):23–34.
18. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830–7.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work was presented as an oral abstract at the 2018 Association of American Medical Colleges Learn Serve Lead meeting.