EDITORIAL

Fool’s gold and chasing unicorns: USMLE Step 1 has no clothes!

Heidi L. Lujan and Stephen E. DiCarlo
Department of Physiology, Wayne State University School of Medicine, Detroit, Michigan

Submitted 30 January 2017; accepted in final form 27 February 2017

Everyone is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.—Albert Einstein

HANS CHRISTIAN ANDERSEN tells the tale of an emperor who loved wearing fine garments and consumed all of his peoples’ hard-earned money on them. He had a different ensemble for each hour of every day and was, by all accounts, the finest dressed man in the land.

One day, two crooks claiming to be weavers convinced the emperor that they were capable of making the finest, lightest, most magnificent fashions the world has ever seen using cloth that was invisible to anyone who was inept or unwise. Impressed, the emperor foolishly paid the swindlers to weave an outfit using their extraordinary cloth.

When the garment was finished, a procession was arranged to show off the emperor’s new apparel, and the entire city gathered. Dressed in his new outfit and believing he looked wonderful, he paraded before his people.

His people believing that the new attire was invisible to anyone who was inept or unwise, offered thunderous applause. None, except a child, was willing to admit that, “The Emperor has no clothes!” Similarly, the United States Medical Licensing Examination (USMLE) Step 1 has “no clothes,” since many people believe that its validity for postgraduate medical residency selection decisions is based on empirical data; however, this belief, like the Emperor’s new clothes, is unwise and unseen.

The USMLE score is one of many factors considered by residency programs in selecting applicants. As such, it is a profoundly “high stakes” exam. High-stakes exams are often the only metric that counts in medical school, and many educators are mainly concerned with test outcomes. The higher the stakes, the more students and professors focus on teaching and learning for the exam. As a result, what is not tested on Step 1 or topics and skills that cannot be tested with a multiple-choice format are not taught, and instruction starts to look like the exam. Furthermore, understandably, students only want to prepare for material that will be tested on Step 1 because of its importance for residency decisions. Of course, this neglects many life skills, including critical thinking, problem solving, communication, interpersonal skills, and compassion. It also creates a competitive environment, a negative attitude toward teaching and learning, and disconnects us from each other and the world around us. This is unfortunate, because it is well documented that patients with physicians who communicate well are more adherent to therapies, more satisfied with care, and less likely to file malpractice suits (4, 7, 15–17, 19). Patients mainly want a doctor who will communicate with them (9). Fortunately, communication skills and other life skills can be taught and evaluated (8).

Furthermore, the scores on USMLE Step 1 are not correlated with reliable measures of medical students’, residents’, or fellows’ clinical skill acquisition (11). In fact, correlations of USMLE Step 1 scores and reliable measures of clinical skills acquisition obtained from nine studies involving 393 students from 2005 to 2010 document that using USMLE Step 1 scores for postgraduate residency selection is neither structured, coherent, nor evidence based (11). This also creates a profound contradiction in that we teach students to be critical thinkers, to use evidence-based medicine and diagnostic tests to guide their decision-making processes; while at the same time relying on a test, to make career changing decisions, that is not evidence based and has little validity for residency selection decisions (12).

Moreover, many students believe that the test defines them such that their intelligence is determined by the grade on a multiple-choice test. However, learning is a continuous process of integrating what is known with new information, and students must understand that knowledge is contextual. That is, information must be tailored and applied to different situations (2). The multiple-choice format of one question, four answers, and only one correct answer creates the illusion of right and wrong, a binary condition that ignores the infinitely fluid nature of information and that knowledge is complex, tentative, evolving, and something that can be improved upon over time. This is disheartening, as it erodes motivation and discourages life-long learning.

The time, stress, and expense associated with Step 1 cannot be overstated, and the social stigma of low scores and possible financial hardships can be overwhelming. As examples, medical students pay approximately $136 million dollars each year in registration fees for the USMLE exams (http://www.usmle.org/performance-data/default.aspx). Tens of millions of dollars are also spent on travel and test preparation expenses. This seems like an unacceptable burden for an exam the holds little validity. Accordingly, it is time to scrutinize our assumptions regarding Step 1 and accept that the value of scores as an objective evaluation of student performance or reliable indicator of student learning is under deep scrutiny and challenge (14).

However, adding a collaborative group testing component to the traditional exam format during years 1 and 2 of the medical school experience may reduce the Step 1 insanity and provide an opportunity for life skills to be taught, practiced, and evaluated (1, 5, 6, 10, 13). Student assessment during the collaborative group testing component of the traditional exams...
during years 1 and 2 can be determined by behavioral, non-cognitive life skills that are based on the observation of the student and the student’s work during the group component. Specifically, communication, collaboration, and problem-solving skills, decision making, and initiative, critical integration of knowledge, professional attitudes, and ethics can all be practiced by the student and evaluated by the faculty during the collaborative group testing component of the traditional exam.

The final grade on the exam can be based on both the performance on the traditional component of the exam and on the behavioral parameters during the group component (10). This approach prevents the entire grade being based on work that rewards only correct answers. This method also rewards students for participation and effort as well as engaging with the material. Accordingly, with the collaborative group testing component, the process of gaining knowledge is as important as the answer. The focus is on social and communication skills and critical thinking and the exchange of ideas (10).

Some students and educators may be concerned that collaborative group testing may bias against students who do not have good communication or social skills. However, everyone has the ability to sharpen their communication and social skills through study and practice. The point is that improving our students’ life skills will result in patients who are more adherent to therapies, more satisfied with care, and less likely to file malpractice suits (4, 7, 15–17, 19). Furthermore, it should be noted that there are efforts to improve adherence to therapies, more satisfied with care, and less likely to file malpractice suits (4, 7, 15–17, 19). Further, our thesis was not a balanced discussion because a preacher does not begin a sermon on the evils of alcohol by wanting them to learn. Which students discover for themselves the concepts and principles we want them to learn.

Adding a collaborative group testing component to the traditional exam format during years 1 and 2 of the medical school experience will not completely change the Step 1 insanity. However, reducing competition and enhancing collaboration and communication with a collaborative group component to the traditional exam is an important step to beginning this process.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

H.L.L. and S.E.D. conceived and designed research; H.L.L. and S.E.D. drafted manuscript; H.L.L. and S.E.D. edited and revised manuscript; H.L.L. and S.E.D. approved final version of manuscript.

REFERENCES

1. Cortright RN, Collins HL, Rodenbaugh DW, DiCarlo SE. Student retention of course content is improved by collaborative-group testing.