

January 2018

## **Learning Outcomes and Performance in Medical School: Programmatic Assessment at NYU School of Medicine**

Colleen Gillespie PHD

*Institute for Innovations in Medical Education*

Sondra Zabar MD

*NYU School of Medicine Department of Medicine*

Adina Kalet MD, MPH

Follow this and additional works at: [https://digitalcommons.nyls.edu/nyls\\_law\\_review](https://digitalcommons.nyls.edu/nyls_law_review)



Part of the [Legal Profession Commons](#), [Legal Remedies Commons](#), and the [Medical Jurisprudence Commons](#)

### **Recommended Citation**

Colleen Gillespie PHD, Sondra Zabar MD & Adina Kalet MD, MPH, *Learning Outcomes and Performance in Medical School: Programmatic Assessment at NYU School of Medicine*, 62 N.Y.L. SCH. L. REV. 73 (2017-2018).

This Article is brought to you for free and open access by DigitalCommons@NYLS. It has been accepted for inclusion in NYLS Law Review by an authorized editor of DigitalCommons@NYLS.

COLLEEN GILLESPIE, PHD, SONDRA ZABAR, MD, AND ADINA  
KALET, MD, MPH

## Learning Outcomes and Performance in Medical School: Programmatic Assessment at NYU School of Medicine

62 N.Y.L. SCH. L. REV. 73 (2017–2018)

**ABOUT THE AUTHORS:** Colleen Gillespie, PhD, Director of Division of Education Quality, Institute for Innovations in Medical Education (IIME), Director of Evaluation and Assessment, Program for Medical Education Innovations and Research (PrMEIR). Sondra Zabar, MD, Professor of Medicine, Director of the Division of General Internal Medicine and Clinical Innovation, NYU School of Medicine Department of Medicine, Co-Director PrMEIR. Adina Kalet, MD, MPH, Director of Research on Medical Education Outcomes (ROMEEO), Co-Director PrMEIR, Professor of Medicine and Surgery. The authors are the founders of PrMEIR, a program in medical education, innovation, and research. The mission of PrMEIR is to advance medical education scholarship and support best practices in medical education through research, faculty development, and an internal grants program. Doctors Gillespie, Zabar, and Kalet work together in ROMEEO, an interdisciplinary research group, to build the evidence base for medical education by establishing linkages between education efforts and patient outcomes. They would like to acknowledge the contributions and support of: the members of PrMEIR, ROMEEO, IIME; the NYU School of Medicine Offices of Medical Education, Student Affairs, and Graduate Medical Education; and the New York Simulation Center.

## I. INTRODUCTION

Although they rarely have formal opportunities to dialogue with one another, medical and legal educators share many assessment challenges. Both professions are involved in significant efforts to better articulate desired learning outcomes and core competencies, which in turn drive innovation in curriculum and training and inform graduation and remediation decisions. In particular, educators from both professions have recognized the need to move well beyond knowledge and its application as the sole-intended outcomes of curricular efforts to help students develop the professions' fundamental skills, including communication and professionalism. Medical and legal educators, therefore, have created detailed rubrics cross-referencing lists of behavioral competencies against various types of competence measurements and assessment rating scales to guide assessment of students.

On a pragmatic level, the authors of this article have struggled with implementation questions: Are our assessments of sufficient quality (that is, are they reliable and valid)? Who completes rating scales, how often, and in what settings? How do we train and support the faculty responsible for using these assessment instruments to ensure the quality judgments made about the competence of our students? What do we do with the data? Who sees it, interprets it, and makes judgments based on it? How do students learn from these assessments and how do students monitor and enhance that learning using assessment data?

In seeking meaningful ways to ensure our graduates are adequately prepared to practice their profession, we find ourselves repeatedly circling back to the fundamental questions. What does it mean to be a competent lawyer or physician? How does competence develop? How individualized and therefore unpredictable is that development? What is the right approach to guide the student who is not yet competent? To what degree does competence assessed in exam settings, ranging from multiple-choice knowledge exams to vivid, realistic, and compelling high-fidelity simulations, predict competence in practice settings? How do we set thresholds and standards for graduation? How do we know when someone is ready for unsupervised practice? How important is assessment of work performance in clinical settings? How do we support learning through assessment and feedback? What is the role of professional identity in helping foster the competencies, values, and practices required of lawyers and doctors? Are there core attributes at the center of our professions that cannot be reduced to a defined set of competencies? How do we know when we can trust our students to practice independently or how much supervision they need to practice safely? Despite having all these fundamental questions and challenges in common, medical and legal educators have few venues in which to share ideas and experiences.

It is in the spirit of sharing that we write the following selective overview of the background and current, broad trends in assessment and professional development in medical education. We describe, as an example, how we are creating an innovative program of assessment for our students at NYU School of Medicine, including our flagship Standardized Patient simulation program that provides the authentic,

developmental, and longitudinal formative and summative assessments of learning outcomes that serve as the foundation for programmatic assessment.

Part II of this paper describes the various movements that have culminated in widespread appreciation for and adoption of comprehensive programs of assessment in medical schools. These include calls for greater accountability, new standards for licensing and certification, the apprenticeship model (wherein clinical supervisors and trainees care for patients, making supervisors dependent upon trainee competence), and the rapidly growing focus on developing the evidence base for effective medical education. Part III describes specific approaches to medical education assessment, including a focus on outcomes, Competency-Based Medical Education (CBME), Entrustable Professional Activities (EPAs), and professionalism and professional identity development. Parts IV and V describe our own efforts to implement a comprehensive program of assessment within NYU's medical school and provide illustrations and examples of many of our specific assessment activities as well as the guiding principles that underpin those activities. Part VI concludes this paper.

## II. BACKGROUND

### A. *Prevailing Forces Supporting Comprehensive Programs of Assessment*

#### 1. *Accountability and Assessment*

Over the last twenty years, the medical profession has experienced serious threats to its social contract. These threats, felt all over the world,<sup>1</sup> are particularly stark in the United States. Challenges stem from the dramatic rise in the financial costs of healthcare, evidence of relatively unsafe and poor quality care,<sup>2</sup> and persistent social inequities in access to and outcomes of care. There has also been a dramatic rise in the cost of a medical education.<sup>3</sup> These factors have led to repeated calls for accountability and return on investment from many sectors, including the public, for the over \$15 billion annual tax dollars that support physician training in the United States.<sup>4</sup>

The emphasis on accountability coupled with significant discoveries in the learning sciences has enabled rapid innovation in medical education assessment.<sup>5</sup> This innovation has been further facilitated by the relatively monolithic nature of medical education policymaking. The Liaison Committee on Medical Education

- 
1. See Richard L. Cruess et al., *Reframing Medical Education to Support Professional Identity Formation*, 89 ACAD. MED. 1446, 1450 (2014) (discussing how, as medicine and societies around the world change, the “details of the social contract change and the nature of the good physician is continuously renegotiated”).
  2. See COMM. ON QUALITY OF HEALTH CARE IN AM., INST. OF MED., *TO ERR IS HUMAN: BUILDING A SAFER HEALTH SYSTEM* 1–3 (Linda T. Kohn et al. eds., 2000).
  3. Louis W. Sullivan, *The Outrageous Cost of Working in Medicine*, WASH. POST, June 9, 2014, <https://www.washingtonpost.com/news/on-leadership/wp/2014/06/09/the-outrageous-cost-of-working-in-medicine/>.
  4. COMM. ON THE GOVERNANCE & FIN. OF GRADUATE MED. EDUC., INST. OF MED., *GRADUATE MEDICAL EDUCATION THAT MEETS THE NATION'S HEALTH NEEDS* 61–62, 98–101 (Jill Eden et al. eds., 2014).
  5. See Syed Amin Tabish, *Assessment Methods in Medical Education*, INT'L J. HEALTH SCIS., July 2008, at 3 (Saudi Arabia).

## LEARNING OUTCOMES AND PERFORMANCE IN MEDICAL SCHOOL

(LCME) is the sole accreditor of North American allopathic medical schools.<sup>6</sup> Further, the American College of Graduate Medical Education (ACGME) maintains a great deal of central control over accreditation of residency training programs.<sup>7</sup> Specialty boards are coordinated by the American Board of Medical Specialties.<sup>8</sup>

One of the main outgrowths of these calls for accountability was a focus on delineating the core set of competencies that define physician practice, which can then be assessed and used to determine if medical schools are producing graduates properly equipped to become effective physicians.<sup>9</sup> Fortunately, there has been a rapidly growing consensus on medical competency frameworks worldwide. In 1998, the ACGME initiated the Outcome Project, which delineated six core competency areas (see Figure 1) and recently mandated the reporting of outcomes in these domains.<sup>10</sup> Almost simultaneously, similar frameworks were proposed in Canada<sup>11</sup> and parts of the United Kingdom,<sup>12</sup> soon followed by Saudi Arabia<sup>13</sup> and many Asian countries.<sup>14</sup> The frameworks are not identical, varying in language, focus (domains of competence versus core roles and functions), and specifics related to contextually or culturally determined particulars. But there are many similarities, and it is fair to say there is now an international consensus on the nature of physician competency.

- 
6. *Scope and Purpose of Accreditation*, LIAISON COMMITTEE ON MED. EDUC., <http://lcme.org/about/> (last visited Jan. 30, 2018). The LCME is sponsored by the AAMC and the AMA. *Relationship with Sponsors*, LIAISON COMMITTEE ON MED. EDUC., <http://lcme.org/about/sponsors/> (last visited Jan. 30, 2018). The two types of licensed doctors in the United States are Doctors of Medicine (MDs) and Doctors of Osteopathic Medicine (DOs). PATRICK WU & JONATHAN SIU, AM. ASS'N OF COLLS. OF OSTEOPATHIC MED., *A BRIEF GUIDE TO OSTEOPATHIC MEDICINE 3* (2d ed. 2015). MDs and DOs are licensed to practice medicine, perform surgery, and prescribe medication; however, MDs focus on allopathic medicine, and DOs take a holistic approach, focusing less on prescription medicine and more on manual manipulation of the body. *Id.* The Commission on Osteopathic College Accreditation accredits DO-granting schools, and the LCME accredits MD-granting schools. *Id.*
  7. *See About Us*, ACCREDITATION COUNCIL FOR GRADUATE MED. EDUC., <http://www.acgme.org/About-Us/Overview> (last visited Jan. 30, 2018).
  8. *See About ABMS*, AM. BOARD MED. SPECIALITIES, <http://www.abms.org/about-abms> (last visited Jan. 30, 2018).
  9. Susan R. Swing, *The ACGME Outcome Project: Retrospective and Prospective*, 29 MED. TCHR. 648, 648–51 (2007) (U.K.).
  10. *Id.*
  11. *See About CanMEDS*, ROYAL COLL. PHYSICIANS & SURGEONS CAN., <http://www.royalcollege.ca/rcsite/canmeds/about-canmeds-e> (last visited Jan. 30, 2018).
  12. *See SCOTTISH DEANS' MED. EDUC. GRP., THE SCOTTISH DOCTOR: LEARNING OUTCOMES FOR THE MEDICAL UNDERGRADUATE IN SCOTLAND 3–4* (3d ed. 2008), <http://www.scottishdoctor.org/resources/scotdoc3.pdf>.
  13. Waleed Hamad Al BuAli et al., *A Framework for a Competency Based Medical Curriculum in Saudi Arabia*, 25 MATERIA SOCIO-MEDICA 148, 148–52 (2013) (Bosn. & Herz.).
  14. Ming-Jung Ho et al., *Does One Size Fit All? Building a Framework for Medical Professionalism*, 86 ACAD. MED. 1407 (2011).

Figure 1: Physician Competency Frameworks

United States (ACGME)	Canada (CanMeds)	United Kingdom (Good Medical Practice)
<ul style="list-style-type: none"> <li>• Medical knowledge</li> <li>• Patient care</li> <li>• Practice-based learning and improvement</li> <li>• Interpersonal and communication skills</li> <li>• Professionalism</li> <li>• Systems-based practice</li> </ul>	<ul style="list-style-type: none"> <li>• Medical expert</li> <li>• Communicator</li> <li>• Collaborator</li> <li>• Manager</li> <li>• Health advocate</li> <li>• Scholar</li> <li>• Professional</li> </ul>	<ul style="list-style-type: none"> <li>• Clinical skills</li> <li>• Practical procedures</li> <li>• Patient investigation</li> <li>• Patient management</li> <li>• Health promotion and disease prevention</li> <li>• Communication</li> <li>• Medical informatics</li> <li>• Basic, social, and clinical sciences and underlying principals</li> <li>• Attitudes, ethical understanding, and legal responsibilities</li> <li>• Decision-making skills and clinical reasoning and judgement</li> <li>• The role of the doctor within health service</li> <li>• Personal development</li> </ul>

The pressure created by accreditation policy and the quickly achieved consensus has stimulated a great deal of innovation and scholarship on the practicalities and best practices for assessing competency. Literature on medical education assessment research is large and growing, and guides and supports changes in policy and practice with significant centers of excellence<sup>15</sup> throughout the United States, Canada, the Netherlands,<sup>16</sup> and the United Kingdom.<sup>17</sup>

15. Centers of excellence are teams or entities that provide guidance or training or both to improve performance in a focus area. See generally Deborah Tolmach Sugarman, *Centers of Excellence*, JAMA NETWORK (Sept. 4, 2013), <http://jamanetwork.com/journals/jama/fullarticle/1734706> (discussing centers of excellence in general and in the medical field); *Why is a 'Center of Excellence' Different from an Institute?*, ADVISORY BOARD: GROWTH CHANNEL (Sept. 14, 2011, 12:00 AM), <https://www.advisory.com/research/market-innovation-center/the-growth-channel/09/what-is-the-difference-between-a-center-of-excellence-and-an-institute> (comparing centers of excellence to institutes).

16. See Cees van der Vleuten, Commentary, *Research in Medical Education: Doctoral Dissertation Reports*, 32 MED. TCHR. 288 (2010) (U.K.).

17. See Geoff Norman, *Research in Medical Education: Three Decades of Progress*, 324 BMJ 1560 (2002) (U.K.), <http://www.bmj.com/content/bmj/324/7353/1560.full.pdf>.

## 2. *Licensing and Certification*

A number of additional factors enable us to refine a competency assessment system. As compared to other professions, the paths that the vast majority of our graduates take upon graduation from medical school, which produces an undifferentiated physician, are relatively well prescribed by highly standardized post-graduate residency training, where physicians specialize. In the United States, licensing requirements are determined nationally through National Board of Medical Examiners exams and requirements,<sup>18</sup> although the actual legal licensing process is administered at the state level.<sup>19</sup> Similarly, national specialty boards determine certification requirements, which are virtually always accepted by employers of and payers to physicians.<sup>20</sup> As a result, educators can understand and to some extent predict the needs of the delimited paths available to graduates of undergraduate medical school programs. Knowing what graduates will need to be able to do establishes a clear mandate for a systematic and comprehensive method of assessing the competencies necessary for practice.

## 3. *Apprenticeship Model*

As medical educators, many of us literally live with our students and graduates. Most medical school faculty practice medicine within the domain of the medical school where they teach.<sup>21</sup> And so, when our medical students enter clinical training, we are taking care of patients side-by-side, with them and through them. This interdependence provides us with a personal investment in their competence, which helps motivate faculty members to engage in meaningful assessment activities.

## 4. *Medical Education Research: Developing Evidence for Education*

Patients, our clients, expect us to be accountable for our trainees' education—to demonstrate the effectiveness of our educational programs by ensuring that participants provide high quality, effective, and efficient patient care. Unfortunately, there is little evidence that directly links what we do in the training of physicians to its impact on the public's health. This combination of forces provides motivation and resources for medical education researchers, a driven group to begin with, to ask difficult and thorny questions about how we know whether our educational programs will ensure that our trainees practice in ways that lead to positive health outcomes. Making direct connections between education and health outcomes is facilitated by high quality

18. See *About the NBME*, NAT'L BOARD MED. EXAMINERS, <http://www.nbme.org/about/index.html> (last updated Nov. 13, 2016); Robert Kocher et al., *Doctors Without State Borders: Practicing Across State Lines*, HEALTH AFF.: HEALTH AFF. BLOG (Feb. 18, 2014), <http://www.healthaffairs.org/doi/10.1377/hblog.20140218.036973/full/>.

19. Kocher et al., *supra* note 18.

20. *Steps Toward Initial Certification and MOC*, AM. BOARD MED. SPECIALTIES, <http://www.abms.org/board-certification/steps-toward-initial-certification-and-moc/> (last visited Jan. 30, 2018).

21. See Molly Cooke et al., *American Medical Education 100 Years After the Flexner Report*, 355 NEW ENG. J. MED. 1339, 1340 (2006).

competency assessment measures. Therefore, improving the measurement quality of assessments is a vital part of the medical education research agenda.

### III. CURRENT TRENDS IN MEDICAL EDUCATION ASSESSMENT

#### A. Outcomes-Based Medical Education

A revolution in medical education occurred in 1910, when Abraham Flexner, an American educator funded by the Carnegie Foundation, produced a damning study of the U.S. medical education system (the “Flexner Report”),<sup>22</sup> which led to dramatic changes and the establishment of scientific rigor and educational excellence as the driving forces behind medical education.<sup>23</sup> The response to the Flexner Report radically improved the quality of physicians by changing the structure and content of the curriculum, but this led to the closure of a large number of U.S. medical schools, especially those in rural areas that served underrepresented, minority communities.<sup>24</sup>

We are now in the early phases of a second revolution in medical education, one that was intentionally initiated on the 100th anniversary of the Flexner Report, when the Carnegie Foundation for the Advancement of Teaching’s Preparation for the Professions program released the extended report of a study they funded (the “Carnegie Report”).<sup>25</sup> The authors of the Carnegie Report summarized four medical education challenges:

- (1) standardizing our outcomes and individualizing the curriculum;
- (2) integrating foundational knowledge with clinical experience;

22. ABRAHAM FLEXNER, CARNEGIE FOUND. FOR THE ADV. OF TEACHING, *MEDICAL EDUCATION IN THE UNITED STATES AND CANADA* (1910).

23. See Edward C. Halperin et al., *Abraham Flexner of Kentucky, His Report*, *Medical Education in the United States and Canada, and the Historical Questions Raised by the Report*, 85 *ACAD. MED.* 203, 203, 209–10 (2010).

24. *BEYOND FLEXNER: MEDICAL EDUCATION IN THE TWENTIETH CENTURY* xii–xiii (Barbara Barzansky & Norman Gevitz eds., 1992) (noting the downturn in women’s medical colleges and black medical schools, particularly in rural areas, after the Flexner Report). The Flexner Report led to medical school licensing and accreditation standards such as defined core basic science curriculum and supervised clinical training in a range of disciplines. Lisa D. Forrester et al., *Overview of Allopathic and Osteopathic Undergraduate Medical Education*, in *GUIDEBOOK FOR UNDERGRADUATE MEDICAL EDUCATION ADMINISTRATION* 5, 6 (Martha Chandler et al. eds., 2014). This led to the closure of many proprietary schools that could not meet the new requirements and stay in business. Ann Steinecke & Charles Terrell, *Progress for Whose Future? The Impact of the Flexner Report on Medical Education for Racial and Ethnic Minority Physicians in the United States*, 85 *ACAD. MED.* 237–38 (2010) (providing an in-depth analysis of how the Flexner Report negatively impacted medical institutions serving racial and ethnic minorities in the United States).

25. MOLLY COOKE ET AL., CARNEGIE FOUND. FOR THE ADV. OF TEACHING, *EDUCATING PHYSICIANS: A CALL FOR REFORM OF MEDICAL SCHOOL AND RESIDENCY* (2010). The Carnegie Foundation has conducted similar studies of a number of other professions including a 2007 report focused on lawyers. WILLIAM M. SULLIVAN ET AL., CARNEGIE FOUND. FOR THE ADV. OF TEACHING, *EDUCATING LAWYERS: PREPARATION FOR THE PROFESSION OF LAW* (2007).



## LEARNING OUTCOMES AND PERFORMANCE IN MEDICAL SCHOOL

- (3) instilling habits of inquiry and improvement in our students; and
- (4) focusing on professional identity formation instead of behaviors or character traits.<sup>26</sup>

This report and the influences leading to it reflect a pivot away from emphasizing curriculum (what we teach and how), toward assessment (what is learned), as the key process of medical education. Rather than struggle to standardize the curriculum, we now measure outcomes so that we can individualize instruction while striving to ensure that all our graduates achieve a uniform high level of competence. A mature implementation of this mastery learning strategy, referred to as progress testing, exists in the Netherlands, where all medical students take a final knowledge exam four times a year for the six years of medical school.<sup>27</sup> Overall scores from these exams, which typically start low and progress on a curve toward a perfect score, along with detailed feedback are shared with students to allow them to monitor their own progress and adjust their study strategies.<sup>28</sup> In addition, educators evaluate the curriculum based on this information.<sup>29</sup>

Medical schools are seeking to address the challenges identified in the Carnegie Report, as evidenced by a greater investment in simulation centers and educational informatics units as core elements for managing and making the best use of the dramatic increase in assessment information.<sup>30</sup> Other evidence of response to the report includes innovation in instruction toward self-assessment and reflection-based activities, such as team-based learning and portfolio-based assessments, where learners are asked to monitor and self-direct their progress based on a range of formative and summative assessments. In addition, there has been development and implementation of workplace-based assessment, designed to capture the full complexity of integrated skills in actual clinical practice contexts,<sup>31</sup> and new focus on measuring medical professional identity development at medical schools around the world.<sup>32</sup> Fully implementing these approaches is quickly leading to an appreciation

---

26. David M. Irby et al., *Calls for Reform of Medical Education by the Carnegie Foundation for the Advancement of Teaching: 1910 and 2010*, 85 *ACAD. MED.* 220, 224–26 (2010).

27. C.P.M. van der Vleuten et al., *Cross Institutional Collaboration in Assessment: A Case on Progress Testing*, 26 *MED. TCHR.* 719, 720 (2004) (U.K.).

28. See René A. Tio et al., *The Progress Test of Medicine: The Dutch Experience*, 5 *PERSP. ON MED. EDUC.* 51, 53–54 (2016) (Neth.).

29. Van der Vleuten et al., *supra* note 27, at 724.

30. Kieran Walsh, Letter to the Editor, *The Future of Simulation in Medical Education*, 29 *J. BIOMEDICAL RES.* 259, 259–60 (2015) (China).

31. *Workplace Based Assessment: A Guide for Implementation*, *GEN. MED. COUNCIL* (Apr. 2010) (U.K.), [http://www.gmc-uk.org/Workplace\\_Based\\_Assessment\\_\\_\\_A\\_guide\\_for\\_implementation\\_0410.pdf\\_48905168.pdf](http://www.gmc-uk.org/Workplace_Based_Assessment___A_guide_for_implementation_0410.pdf_48905168.pdf) (last visited Jan. 30, 2018).

32. Adina Kalet et al., *Measuring Professional Identity Formation Early in Medical School*, 39 *MED. TCHR.* 255, 255 (2016) (U.K.).

for the need to embed all these activities within a well-planned, comprehensive, longitudinal “program” of assessment.

### B. Competency-Based Medical Education (CBME)

Should medical education, or any education for that matter, be time based any longer? Focus on meaningful measurement of outcomes of medical education, along with meticulous defining of competence, has coalesced into what some call CBME. This movement challenges the idea that “dwell time”—spending the requisite four years (in the United States; six in Europe) in training—is what makes a competent physician.<sup>33</sup> In fact, evidence suggests that some people need more time while others need less to achieve the same level of performance.<sup>34</sup> In a CBME framework, our job is to define final outcomes, measure individuals’ abilities, give feedback and tailored “just in time” learning,<sup>35</sup> and provide opportunities for further practice until trainees achieve the desired state. Only then are trainees promoted. This approach is learner centered, not course or curriculum based. It requires valid measurement of competence and the establishment of mastery standards based on the ultimate outcome.

These are the outcomes the ACGME was aiming at in 2005, and over the years the competencies have been defined in great detail. Milestones—developmental standards by stages of training—have been identified and implemented in residency training.<sup>36</sup> Hundreds of distinct milestones for each specialty have been defined, and systems have been put in place to measure, monitor, and report them to the ACGME.<sup>37</sup> For example, Internal Medicine Residency Programs, through a consensus project supported by the ACGME and the American Board of Internal Medicine, initially defined more than 120 milestones ranging from “Gathers and Synthesizes Essential and Accurate Information To Define Each Patient’s Clinical Problem(s),” to “Demonstrates Appropriate Utilization and Completion of Health Records.”<sup>38</sup>

Recognizing the impracticality of assessing residents’ developmental progress across so many milestones, the consensus group established twenty-two “reporting”

33. See Mark A. Albanese et al., *Defining Characteristics of Educational Competencies*, 42 MED. EDUC. 248, 250–51 (2008).

34. Carol L. Carraccio et al., *From the Educational Bench to the Clinical Bedside: Translating the Dreyfus Developmental Model to the Learning of Clinical Skills*, 83 ACAD. MED. 761, 766 (2008).

35. Just in time learning is an instructional strategy aimed at maximizing classroom time by gearing students toward identifying learning needs or gaps, for example, through self-assessment quizzes, and addressing these needs using targeted materials, such as reading and online resources. See JEROEN J. G. VAN MERRIËNBOER & PAUL A. KIRSCHNER, *TEN STEPS TO COMPLEX LEARNING: A SYSTEMATIC APPROACH TO FOUR-COMPONENT INSTRUCTIONAL DESIGN* 199–200 (2d ed. 2013). Just in time learning is a fundamental feature of evidence-based models of complex learning and eLearning. *Id.* at 206.

36. Eric S. Holmboe et al., *Milestones and Competency-Based Medical Education in Internal Medicine*, 176 JAMA INTERNAL MED. 1601, 1601 (2016).

37. Susan R. Swing et al., *Educational Milestone Development in the First 7 Specialties to Enter the Next Accreditation System*, 5 J. GRADUATE MED. EDUC. 98 (2013).

38. See William Iobst et al., *Internal Medicine Milestones*, 5 J. GRADUATE MED. EDUC. (SUPP.) 14, 15–23 (2013).

milestones and created nine-point rubrics for clinical faculty to use to rate residents every six months based on observations and assessments in multiple clinical contexts. These ratings are then reported to the ACGME to provide a national perspective on milestone achievement within each specialty.<sup>39</sup> Exhausting! This effort has served us well in many ways, but it has at the same time been cumbersome to implement. Despite the commitment and engagement across the profession, there remains an unsatisfying and incomplete attempt to capture what it means to be a physician.<sup>40</sup> From this frustration has come three exciting assessment initiatives in medical education: (1) EPAs,<sup>41</sup> (2) Professional Identity Formation (PIF),<sup>42</sup> and (3) “programs of assessment *for learning*.”<sup>43</sup>

### C. Entrustable Professional Activities (EPAs)

We are educating professionals to perform complicated patient care tasks and fulfill complex healthcare responsibilities where they must integrate knowledge, skills, and values in challenging clinical contexts to consistently achieve near-perfect levels of performance. This is a tall order. As previously discussed, competencies are “person descriptors” of individuals’ knowledge, skills, attitudes, and values, broken down into component parts.<sup>44</sup> Measures of competence, even when performed at optimal levels, do not necessarily indicate that we can trust a student to perform the activities of an independent practitioner in authentic settings. Students may be *unable to or choose not to* “transfer” their competence to the complexities of actual practice demands.

EPAs are descriptions of the work trustworthy physicians must do. For example, a physician might gather a patient’s history, conduct a physical examination, perform a lumbar puncture, request consent for an autopsy, deliver bad news, and arrive at a likely diagnosis.<sup>45</sup> EPAs have been proposed as orienting end goals for medical training because they appeal to physicians because, as compared to the core competencies that drove us over the past few decades, they are more realistic, less

39. *Milestones*, ACCREDITATION COUNCIL FOR GRADUATE MED. EDUC., <http://www.acgme.org/What-We-Do/Accreditation/Milestones/Overview> (last visited Jan. 30, 2018).

40. See Carol Carraccio et al., Letter to the Editor, *Milestones: Not Millstones but Stepping Stones*, 6 J. GRADUATE MED. EDUC. 589, 589 (2014).

41. See generally ASS’N OF AM. MED. COLLS., CORE ENTRUSTABLE PROFESSIONAL ACTIVITIES FOR ENTERING RESIDENCY (2014), <https://members.aamc.org/eweb/upload/core%20epa%20curriculum%20dev%20guide.pdf> (providing a guide for medical curriculum developers by delineating thirteen EPAs that all entering residents must be able to perform without supervision).

42. See generally Kalet et al., *supra* note 32, at 255–60 (presenting an overview of PIF use in a professional curriculum).

43. See generally Lambert W.T. Schuwirth & Cees P.M. van der Vleuten, *Programmatic Assessment: From Assessment of Learning to Assessment for Learning*, 33 MED. TCHR. 478 (2011) (detailing the shift of assessment’s role in education from assessment of learning to assessment for education) (U.K.).

44. ORG. FOR ECON. CO-OPERATION & DEV., THE DEFINITION AND SELECTION OF KEY COMPETENCIES 4 (2005) (Fr.), <https://www.oecd.org/pisa/35070367.pdf>.

45. See Olle ten Cate, *Entrustability of Professional Activities and Competency-Based Training*, 39 MED. EDUC. 1176, 1177 (2005).

abstract, and capture something about the expertise required to “put it all together” to function effectively in a real setting.<sup>46</sup>

As an approach to defining and measuring medical competence, EPAs have been embraced by the Association of American Medical Colleges (AAMC), the national membership organization that serves to support medical education and training.<sup>47</sup> In 2014, the AAMC released a set of thirteen core EPAs for graduation from medical school that were established through an extensive consensus process.<sup>48</sup> These core EPAs were proposed for use by all medical schools as graduation “endpoints” to ensure that the residency programs receiving these newly graduated “physicians in training” would know what to expect on the first day of residency.<sup>49</sup> As a pilot program, the AAMC invited ten medical schools to explore how curricular and assessment strategies should align with these core EPAs.<sup>50</sup> The EPAs are likely to become a required framework for outcome assessment in the near future.<sup>51</sup> In some ways, this is simply (though not easily) a reshuffling of the work done to define core competencies and developmental milestones into different, more authentic groupings. However, the new and appealing piece is the introduction of the idea of “entrustability,” which incorporates elements of trust, trustworthiness, and entrustment decisions into assessment.<sup>52</sup>

Medical education has always been based on a model of supporting learners in taking on these responsibilities at appropriately increasing levels of independence. The faculty supervising trainees in performing patient care tasks must constantly make judgments about the degree to which they can “trust” the trainee to perform safely. The faculty member must distinguish if the trainee: (1) can do the task at all (if not, the trainee should simply observe), (2) can be trusted to do elements of the task under close observation, (3) can be trusted to perform the task and instructed to call for immediate help if needed, or (4) can be trusted to perform the task without supervision.<sup>53</sup> Since trust is the basis for the judgments clinical supervisors make in

46. See H. Carrie Chen et al., *The Case for Use of Entrustable Professional Activities in Undergraduate Medical Education*, 90 ACAD. MED. 431, 433 (2015); Claire Touchie & Olle ten Cate, *The Promise, Perils, Problems and Progress of Competency-Based Medical Education*, 50 MED. EDUC. 93, 96 (2016).

47. See Kim Krisberg, *Competency-Based Education Improves Transition from Medical School to Residency*, ASS’N AM. MED. COLLS.: AAMCNEWS (Sept. 27, 2016), <https://news.aamc.org/medical-education/article/competency-based-education-residency/>.

48. See ASS’N OF AM. MED. COLLS., *supra* note 41.

49. See *Milestones*, *supra* note 39.

50. *The Core Entrustable Professional Activities (EPAs) for Entering Residency*, ASS’N AM. MED. COLLS., <https://www.aamc.org/initiatives/coreepas/> (last visited Jan. 30, 2018).

51. See *Milestones*, *supra* note 39; see also Kimberly Lomis et al., *Implementing an Entrustable Professional Activities Framework in Undergraduate Medical Education: Early Lessons from the AAMC Core Entrustable Professional Activities for Entering Residency Pilot*, 92 ACAD. MED. 765 (2017).

52. See Janelle Rekman et al., *Entrustability Scales: Outlining Their Usefulness for Competency-Based Clinical Assessment*, 91 ACAD. MED. 186 (2016).

53. See Olle ten Cate, *Nuts and Bolts of Entrustable Professional Activities*, 5 J. GRADUATE MED. EDUC. 157, 158 (2013).

allowing trainees increasing independence, the most interesting conversations in the current medical education literature unpack how these “entrustment” decisions are and should be made.<sup>54</sup> This requires understanding the bases and biases of these expert judgments, sometimes referred to as “rater cognition” in the assessment literature.<sup>55</sup> These expert judgments tend to be highly internally reliable (our judgments are consistent with our own judgments), but also idiosyncratic (we tend not to agree with other experts), which highlights the importance of understanding the discrepancies among competence judgments.<sup>56</sup> This presents a challenge to incorporating assessment of EPAs into medical schools: How do we achieve the best balance of the strengths and weaknesses of expert judgment?

The emerging consensus is that we must dramatically expand the number of observations.<sup>57</sup> Therefore, assessing the “entrustability” of medical trainees should be based on many samples of the “professional activity”—requiring progressively more integration of constituent competencies and coming from increasingly realistic settings and situations (from simulation to actual clinical practice, for example)—based on the judgment of many different experts. All this suggests that we need to know more about the developmental nature of clinical competence and entrustability.

Figure 2 shows an idealized curve of the development of a medical trainee’s clinical skills plotted along with theoretical cut-offs for entrustment decisionmaking. The skills development curve is predictably steep at certain stages (usually in the novice phase), and then levels off, reflecting that additional gains in competence require increasing levels of effort over longer times. Not everyone develops at the same pace. Achieving the ultimate goal state is not guaranteed. It requires persistence,

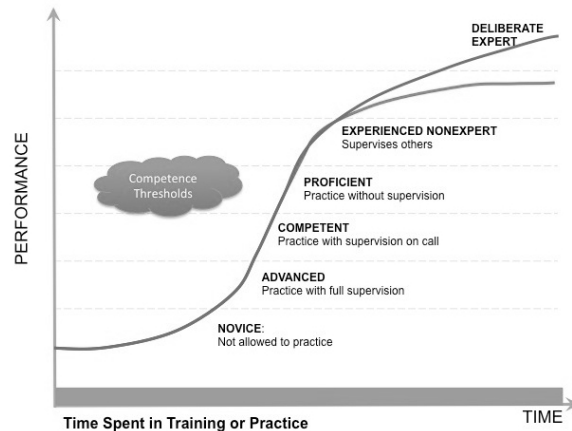


Figure 2: Skill Development and Entrustment

54. See, e.g., Andrea Gingerich et al., *Rater-Based Assessments as Social Judgments: Rethinking the Etiology of Rater Errors*, 86 *ACAD. MED. (SUPP.)* S1 (2011); Andrea Gingerich et al., *Seeing the ‘Black Box’ Differently: Assessor Cognition from Three Research Perspectives*, 48 *MED. EDUC.* 1055 (2014) [hereinafter Gingerich et al., *Assessor Cognition*].
55. See Qie Han, *Rater Cognition in L2 Speaking Assessment: A Review of the Literature*, 16 *TCHR. C. COLUM. U. WORKING PAPERS IN APPLIED LINGUISTICS & TESOL* 1, 3 (2016); Rekman et al., *supra* note 52, at 186–88.
56. See Rekman et al., *supra* note 52.
57. See Brian Hodges, *Assessment in the Post-Psychometric Era: Learning to Love the Subjective and Collective*, 35 *MED. TCHR.* 564, 565 (2013) (U.K.); see also Lambert Schuwirth & Julie Ash, *Assessing Tomorrow’s Learners: In Competency-Based Education Only a Radically Different Holistic Method of Assessment Will Work. Six Things We Could Forget*, 35 *MED. TCHR.* 555 (2013) (U.K.); Schuwirth & Van der Vleuten, *supra* note 43.

grit, and a willingness to spend time in frequent cycles of practice with feedback. This behavior is called deliberative practice and is required for the development of expertise in many professions and activities that are characterized as optimal performance domains, like competitive athletics and chess.<sup>58</sup>

*D. Professional Identity Formation (PIF): What Makes Medical and Law Schools Professional?*

We have heard our higher education colleagues refer to us somewhat disparagingly as trade schools, turning out uniformly certified providers of services. This is no doubt partly due to our high-profile engagement in and increasing emphasis on detailed outcome and competency-based and measurement-focused assessment frameworks. In fact, we have also heard such talk in our own curriculum committees as medical school faculty express their fears about the consequences of competency-based professional education. We acknowledge that there will likely be unintended negative consequences of becoming so explicit about the product of our education system. But we believe we have a significant contribution to make to general higher education programs in two distinct areas.

First, we are turning away from a focus on assessment *of* learning, which provides only final grades to ensure short-term retention of learning, toward a focus on assessment *for* learning,<sup>59</sup> which harnesses the power of a program of assessment in which many frequent “pieces” of assessment information are collected across contexts and over time and through the judgment of many experts.<sup>60</sup> These pieces of information are deliberately curated to compel deeper, more durable learning with greater retention and more expert application of that learning in highly complex situations.<sup>61</sup> The following section describes one example of how we are approaching the development of a rich program of assessment for learning at NYU School of Medicine.

Second, we are increasingly teaching toward the creation of a unique professional identity. By doing so, we must explicitly assess, consider, and address the moral, ethical, psychological, and personal behavioral expectations of our students. Beyond simply describing the virtues, character, and behaviors of the ideal professional, we must judge the nature of our students’ developing medical professional identity and socialization into the profession. At NYU School of Medicine, we have been exploring how this assessment of the formation of professional identity should be incorporated as a critical component of our program of assessment alongside medical knowledge, integrated clinical skills, and health systems science. Understanding how all the components of competence interrelate and integrate with the development of a sound

58. K. Anders Ericsson, *Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains*, 79 ACAD. MED. (SUPP. 10) S70, S72–73 (2004).

59. Schuwirth & Van der Vleuten, *supra* note 43, at 478.

60. See ALEXANDER W. ASTIN ET AL., AM. ASS’N OF HIGHER EDUC., PRINCIPLES OF GOOD PRACTICE FOR ASSESSING STUDENT LEARNING (1992).

61. ORG. FOR ECON. CO-OPERATION & DEV., FORMATIVE ASSESSMENT: IMPROVING LEARNING IN SECONDARY CLASSROOMS 2 (2005) (Fr.), <https://www.oecd.org/edu/ceeri/35661078.pdf>.

professional identify is likely to lead to a richer and more meaningful framework for making entrustment judgments and ensuring the best outcomes for patients.

### IV. BUILDING A PROGRAMMATIC APPROACH TO ASSESSMENT AT NYU

#### A. Overview

NYU School of Medicine, like all medical schools, is continuously striving to ensure that its graduates are fully prepared to become effective physicians. The broad forces and specific innovations in medical education discussed previously have led us to consider how best to implement a coordinated, comprehensive program of assessment. Our curriculum focuses on foundational medical knowledge for the first year and a half of medical school, including eleven basic science “modules” and an intensive “doctoring” course that focuses on core communication, interviewing and history-gathering skills, physical examination, and early diagnostic reasoning skills. The acquisition of knowledge is assessed via multiple choice questions on comprehensive examinations that strive to incorporate an equal mix of factual and application questions. The “doctoring” course incorporates frequent skills practice with formative feedback, using Standardized Patients (SPs)—actors trained to play patients—in our simulation center as well as real patients at the bedside. It also includes more summative assessments of the integration of skills at targeted points. After that, students begin “clerkships” in which they rotate through their core clinical experiences, typically spending between four and eight weeks in each of seven specialties of the medical profession.<sup>62</sup> The students then choose specific clinical experiences of interest—such as emergency medicine or radiology—and more advanced rotations where they begin to function almost as residents in what are referred to as Advanced Clerkships, in preparation for residency. Students take several national examinations throughout the process to determine whether they have mastered core medical knowledge, the application of that knowledge to clinical problems, and basic clinical skills. The bulk of assessment data points, however, come from faculty (and resident) ratings of students’ performance across core competencies in the clinical clerkships. Threaded throughout are the more structured opportunities for practice, feedback, and assessment that simulation makes possible, including a high-stakes, pass/fail clinical examination. The examination involves the demonstration and assessment of core communication skills, history gathering, physical examination, and clinical reasoning across eight general medicine scenarios.

These assessments can be described along a continuum moving from knowledge to action that George Miller has depicted as a pyramid (shown in Figure 3),<sup>63</sup> a depiction

---

62. See *Stage Two: Clerkship Year*, NYU LANGONE HEALTH, <http://med.nyu.edu/education/md-degree/md-curriculum/stage-two-clerkship-year> (last visited Jan. 30, 2018). Specialty areas include internal medicine and ambulatory care, pediatrics, obstetrics and gynecology, surgery, psychiatry, and neurology. *Id.*

63. George E. Miller, *The Assessment of Clinical Skills/Competence/Performance*, 65 ACAD. MED. (SEPT. SUPP.) S63, S63 (1990).

that has been widely embraced and modified.<sup>64</sup> A key component of a program of assessment is threading domains through this pyramid structure, showing how initial foundational knowledge and skills are related to development of subsequent skills, establishment of competence in specific domains, and integration of those competencies into the ability to “do” the work of a physician with decreasing supervision.

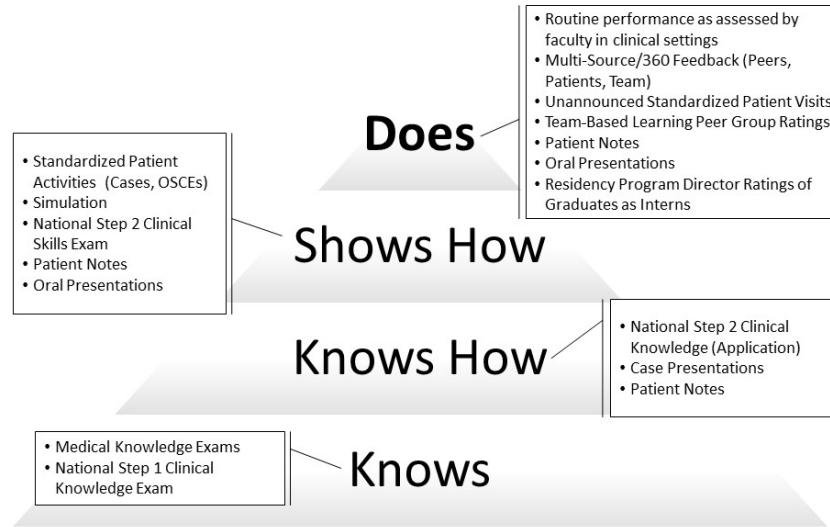


Figure 3: Miller's Pyramid of Competence

### B. Goals of Our Program of Assessment

The main goal of this program is to provide students with a clear roadmap to the process of becoming a physician. This includes constant feedback on their progress toward that goal, focusing particularly on what they need to do to achieve that progress and how the medical school can help them. As a medical school, we also use this program of assessment to continually monitor the quality and effectiveness of our curriculum.

### C. Principles Informing Our Program of Assessment

The following key principles have been used throughout our assessment program and serve as a framework for making decisions about assessment.

- (1) Assessment should be designed for learning. It is important to determine how best to integrate assessment into learning through cycles of feedback, reflection, and deliberate practice.
- (2) Assessment data should be used to facilitate understanding, interpretation, and action. The power of assessment data should

64. Richard L. Cruess et al., *Amending Miller's Pyramid to Include Professional Identity Formation*, 91 ACAD. MED. 180, 180–81 (2016).



## LEARNING OUTCOMES AND PERFORMANCE IN MEDICAL SCHOOL

be “harnessed” through thoughtful use of informatics and data analytics and visualization to provide feedback, support reflection and learning, and individualize education.

- (3) Competence and competent performance in clinically meaningful contexts should be the focus of assessment but should build from an awareness of the foundational knowledge, skills, values, and attitudes that, when fully integrated, form the core of professional responsibilities.
- (4) Assessment of competencies, especially the integration of competencies into professional activities, should be titrated across increasingly realistic and therefore more challenging contexts, from exams to simulation to clinical practice.
- (5) Assessment approaches should be developmental and longitudinal—developmental in terms of focusing on the appropriate level of learning outcome at the appropriate time in the trainees’ trajectory toward mastery and longitudinal in terms of documenting progress toward defined end points.
- (6) Assessment must ultimately permit experts to determine with confidence whether trainees are prepared to perform the essential, synthetic, authentic responsibilities of the profession.

### *D. Examples from NYU’s Program of Assessment*

Below we provide specific examples from our emerging “program” of assessment.<sup>65</sup> Each is meant to illustrate application of the trends in assessment in medical education we have discussed thus far,<sup>66</sup> and the development and implementation of each was informed by the principles described above. We describe in depth our SP Program because these simulation activities serve as a key substrate for our assessment program.

#### *1. Ensuring Structured Feedback*

While students receive feedback throughout their clinical experiences, and our accreditation body, the LCME, mandates mid-clerkship feedback, we determined that it was important for students and their preceptors to have a scheduled, structured, recorded, in-person feedback meeting midway through each clerkship. An assessment form first completed by a student as a self-assessment and then by supervising faculty as an “expert” assessment of the student’s performance to date is accessible through a mobile device—in our case, an iPad (Figure 4). The iPad provides mobility to

65. C.P.M. van der Vleuten et al., *A Model for Programmatic Assessment Fit for Purpose*, 34 MED. TCHR. 205, 205–12 (2012) (U.K.).

66. See discussion *supra* Part III.

### Step 4: Student and Preceptor Discuss

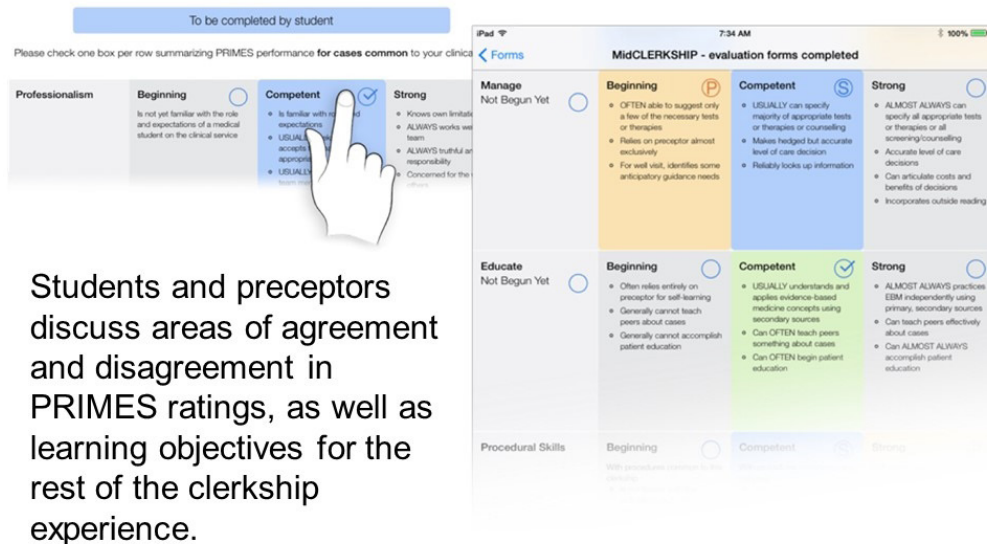


Figure 4: Structured Feedback

facilitate assessment in the workplace (the clinical unit), works without internet access (which is necessary in some of our clinical environments), and requires in-person, face-to-face interaction between student and evaluator.<sup>67</sup> The design of the form and process facilitates comparison between a student's self-assessment and faculty assessment. It is based on a widely accepted framework for the development of clinical competence<sup>68</sup> that can be used across all clinical rotations and over time to document progress. The form and process were explicitly designed to support student reflection and deliberate practice by asking the student, in collaboration with and under the guidance of the evaluator, to identify three learning goals for the remaining half of the clerkship.

#### 2. Assessment to Support Reflection and PIF

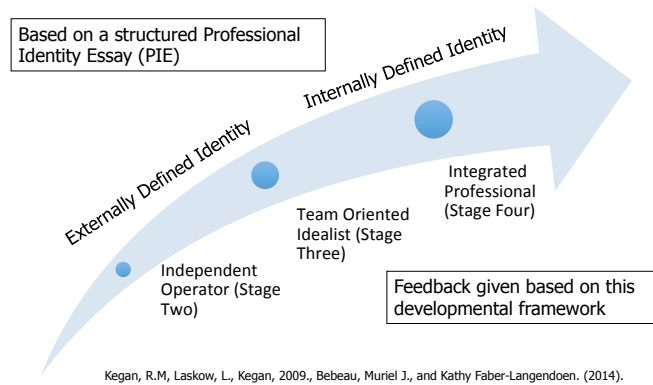
We have developed a series of exercises that require students to reflect on their progress as physicians in training and participate in mentoring and advisement around those reflections. The reflections are scheduled throughout the curriculum to occur at key moments of development and transition. The initial reflection, within the first four months of training, asks students to reflect on preliminary data on their emerging clinical skills and use that data to set goals for subsequent skills

67. Our workflow involves the student physically handing her iPad, containing a completed self-assessment, to the evaluator, who completes her evaluation of the student in real time, and then they discuss in person.

68. Tony Ogburn & Eve Espey, *The R-I-M-E Method for Evaluation of Medical Students on an Obstetrics and Gynecology Clerkship*, 189 AM. J. OBSTETRICS & GYNECOLOGY 666, 669 (2003). This framework is Reporter, Interpreter, Manager, and Educator (RIME) and is based on stages of competency development by student roles. *Id.* at 666.

## LEARNING OUTCOMES AND PERFORMANCE IN MEDICAL SCHOOL

development. This is repeated when students have accumulated sufficient additional information on their clinical skills. As students transition from the pre-clinical stage, focusing on the acquisition of medical knowledge and core skills, to the clinical stage, focusing on applying that knowledge and those skills in practice, they are asked to reflect on their preparedness. Students then, partway through these clinical stages, develop their self-awareness skills through reflection. Finally, as students consider their future training and career pathways, they engage in reflection exercises designed to help them choose their specialization, find and position themselves for the most appropriate residency programs, and prepare for residency. Students are mentored throughout these exercises, with advisors and “coaches” providing feedback on the quality of the reflection, guidance in establishing realistic learning goals, and help in engaging in deliberate practice. In addition, we are piloting an explicit focus on professional identity development and are therefore requiring students to write a structured essay several times throughout their education. The content of that essay can then be assessed to describe the writer’s stage of PIF,<sup>69</sup> and students receive feedback on their stage of development (Figure 5).



**Figure 5: Assessment of Professional Identity**

Students then, partway through these clinical stages, develop their self-awareness skills through reflection. Finally, as students consider their future training and career pathways, they engage in reflection exercises designed to help them choose their specialization, find and position themselves for the most appropriate residency programs, and prepare for residency. Students are mentored throughout these exercises, with advisors and “coaches” providing feedback on the quality of the reflection, guidance in establishing realistic learning goals, and help in engaging in deliberate practice. In addition, we are piloting an explicit focus on professional identity development and are therefore requiring students to write a structured essay several times throughout their education. The content of that essay can then be assessed to describe the writer’s stage of PIF,<sup>69</sup> and students receive feedback on their stage of development (Figure 5).

Finally, we support students by understanding the expectations of the profession and ensuring they are meeting those expectations by assessing explicit dimensions of

**Figure 6: Dashboard of Clerkship Feedback over Time:  
Learner-Centered Displays of Formative Assessment Data**

Preceptor Ratings of Student						
Clerkship	Pediatrics	Surgery	Medicine	Ambulatory Care	Psychiatry	CNS
Rotation Period	1/2/14-2/16/14	2/18/14-3/19/14	3/24/14-4/18/14	4/19/14-5/18/14	5/24/14-6/30/14	7/1/14-8/18/14
Professionalism	Competent	Competent	Competent	Strong	Strong	Strong
Reporter	Beginning	Competent	Strong	Strong	Strong	Strong
Interpreter	Beginning	Beginning	Competent	Competent	Competent	Strong
Manager	Not Begun Yet	Beginning	Beginning	Beginning	Competent	Beginning
Educator	Not Begun Yet	Not Begun Yet	Not Begun Yet	Not Begun Yet	Not Begun Yet	Not Begun Yet
Procedural Skills	Beginning	Beginning	Beginning	Competent	Competent	Competent

69. See ROBERT KEGAN & LISA LASKOW LAHEY, IMMUNITY TO CHANGE 112–15 (2009).

professionalism throughout their clinical clerkships.<sup>70</sup> Students are assessed on these dimensions at the end of each clerkship by one to six of their clinical supervisors (Figure 6). While we have been providing high-level feedback to students as part of their clerkship grade, we plan to provide more granular data on each individual assessment so lapses in “professionalism” can be identified, and consistency in students’ abilities to meet professional standards across contexts with high levels of responsibility and complexity can be increased.

A true “program of assessment” is not possible without effective methods of collecting, managing, and displaying the rich array of data elements. NYU School of Medicine is fortunate to have been an early leader in the field of educational informatics. It is home of the Institute for Innovations in Medical Education, which maintains an education data warehouse and supports a team of programmers and data analysts. These analysts can not only manage large, complex educational data sets and streams, but also develop and implement “dashboard” views of the data to facilitate interpretation and action.<sup>71</sup> As an example, we have taken the mid-clerkship feedback assessments described earlier and displayed them for students as a longitudinal array so that they can monitor the pattern of their assessments over time and track their learning goals across the diverse set of clinical clerkships through which they rotate.

We are developing the prototype for a comprehensive learner-centered dashboard which will house various types of assessments created from multiple sources. It will display those data points in alignment with specific competency domains, over time, and along a continuum that ranges from simulated to real-world settings.<sup>72</sup> The goals of these dashboards are to facilitate reflection and deliberate practice so that students can monitor and take ownership of their progress and to drive curriculum broadly for cohorts of students and specifically for individual or sub-groups of individuals who may need specialized resources. Figure 7 displays the “student portal” to their data organized by domain.

Narrative assessment<sup>73</sup> is currently making a comeback as educators re-recognize the power of the more in-depth and nuanced discussion of students’ strengths and weaknesses that is made possible through open-ended textual assessments.<sup>74</sup> Students

70. Dimensions of professionalism can include demonstrating concern for others, effectively communicating with patients and healthcare team members, being truthful, reliable, and trustworthy, knowing one’s limits and seeking help appropriately, and taking on and fulfilling responsibilities.

71. Inst. for Innovations in Med. Educ., *Institute for Innovations in Medical Education Programs & Divisions*, NYU LANGONE HEALTH, <https://med.nyu.edu/institute-innovations-medical-education/our-programs-divisions> (last visited Jan. 30, 2018).

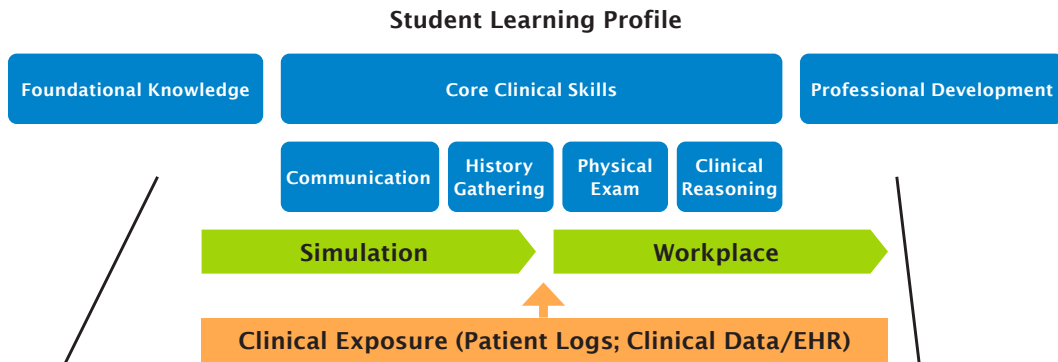
72. The types of assessments might range from exam scores to performance assessments or from formative to summative assessments. The sources from which they are created include peers, faculty, patients, and team members.

73. Narrative assessment consists of replacing or complementing numerical ratings and grades with rich, in-depth written descriptions of student performance across multiple observations. See Gingerich et al., *Assessor Cognition*, *supra* note 54, at 1056.

74. See Janice L. Hanson et al., *Narrative Descriptions Should Replace Grades and Numerical Ratings for Clinical Performance in Medical Education in the United States*, *FRONTIERS PSYCHOL.*, Nov. 21, 2013, at 1, 6–8 (Switz.).

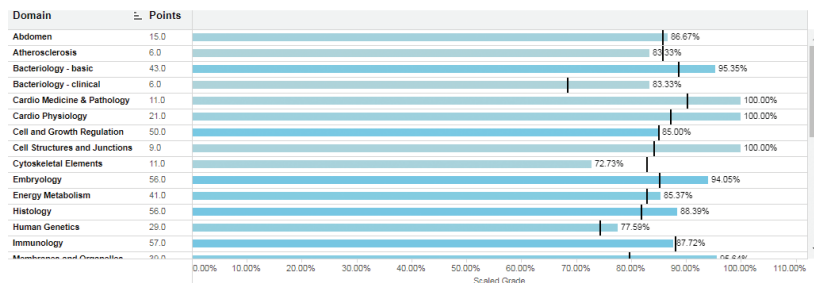
# LEARNING OUTCOMES AND PERFORMANCE IN MEDICAL SCHOOL

Figure 7: Student Dashboard for Learning



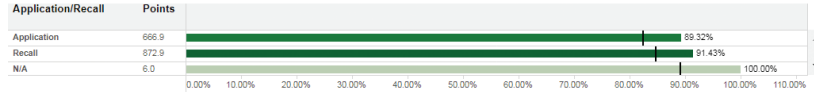
## Foundational Knowledge

### Breakdown by Knowledge Domain

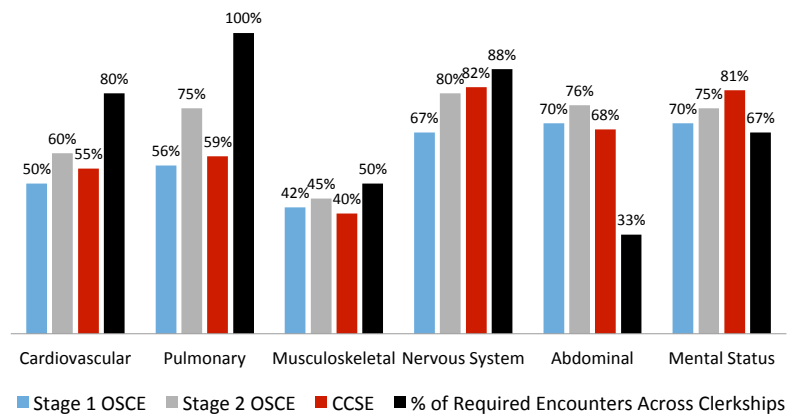


\*(Points available toward this competency/Total points in this report) The number of points will exceed the number of exam questions if any questions are worth more than one point. (E.g., A 100 question exam which contains a five point essay will total 104 points.)

### Breakdown by Application/Recall



## Feedback on Physical Exam Skills



Combine specific feedback data with degree of clinical exposure (integrate performance with curriculum)

appreciate that these narratives tend to provide more information and direction than are generally available from quantitative ratings and feel that such assessments are more personalized to them as individuals. They therefore find narrative assessments more compelling and motivating. However, the challenge with these assessments is that there are often thousands of comments available for any one student, many of them rich and complex—and therefore, not easily reduced, summarized, and organized into displays that facilitate understanding and action. We are experimenting with natural language processing to develop sentiment analysis algorithms that permit initial sorting of comments based on negative and positive indices.

### 3. *Time-Based Versus Competency-Based Progression*

The shift from determining progression based on time in medical school (and of course the completion of required experiences during that time) to competency-based progression is a tremendously challenging goal. Structuring a medical school experience that stops when students have achieved competence does not fit well within current models of scheduling, tuition, and resource allocation. Instead, therefore, medical schools around the world are experimenting with accelerated pathways for selected students.<sup>75</sup> Any attempt to use competence as the threshold for progression, however, depends on quality assessment data—data that is fair, consistent, and accurate and therefore would permit educators to have confidence in promotion decisions made based on those data.

Our current approach to competency-based, mastery assessment at NYU is most mature in the area of communication skills.<sup>76</sup> We have developed a consistent assessment framework for measuring such skills based on a conceptual model of the core functions of communication in medicine, translated that framework into an assessment instrument containing empirically supported, observable behaviors that have been shown to lead to better patient outcomes, and implemented a rating scale of “not done,” “partly done,” and “well done,” which signals that only “well done” skills/behaviors are effective. Using this same assessment instrument throughout medical school allows us to track students as they progress toward mastering core communication

75. See Joan Cangiarella et al., *Three-Year MD Programs: Perspectives from the Consortium of Accelerated Medical Pathway Programs (CAMPP)*, 92 ACAD. MED. 483 (2017).

76. While communication has always been viewed as an essential skill in medical practice, the two organizations that accredit medical schools in North America formally adopted a resolution in 1995, which stated that “[t]here must be specific instruction in communication skills as they relate to physician responsibilities, including communication with patients, families, colleagues and other health professionals.” LIAISON COMM. ON MED. EDUC., FUNCTIONS AND STRUCTURE OF A MEDICAL SCHOOL (2003). Communication skills, initially viewed in a more limited way as consisting mainly of bedside manner or history taking, have come to be seen as a measurable clinical skill. See Gregory Makoul et al., *Essential Elements of Communication in Medical Encounters: The Kalamazoo Consensus Statement*, 76 ACAD. MED. 390, 390–92 (2001). The organization overseeing U.S. residency training programs identifies “interpersonal and communication skills that result in effective information exchange and teaming with patients, their families and other health professionals” as a core area of competency. MARK D. SULLIVAN, THE PATIENT AS AGENT OF HEALTH AND HEALTH CARE 23 (2017); see Paul Batalden et al., *General Competencies and Accreditation in Graduate Medical Education*, 21 HEALTH AFF. 103 (2002).

## LEARNING OUTCOMES AND PERFORMANCE IN MEDICAL SCHOOL

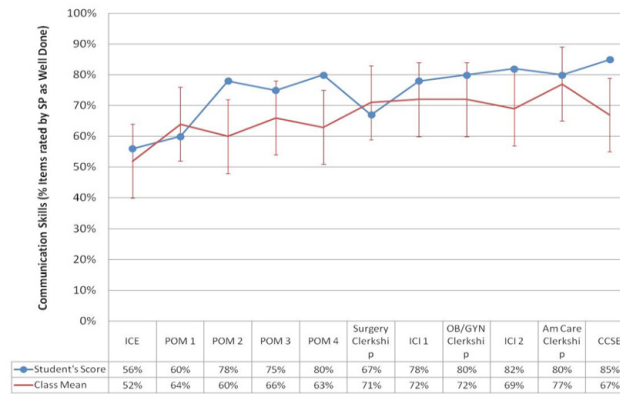
skills. In particular, we have scored this instrument based on the percentage of items for which students received a “well done” rating—this score then, over time, describes students’ mastery of effective communication skills. Figure 8 plots one student’s communication scores throughout medical school—as assessed in Objective Structured Clinical Examinations (OSCEs), a multi-station SP and performance based exam—in comparison to the entire class of students and shows progress toward mastery, even as some of the cases in the clinical examinations become more challenging.

We have found, however, that progress toward mastery is not always an interrupted line trending upwards. As shown in Figure 9, based on a random sample of students’ three communication skill data points, we see that they, on average, are improving. But a review of individual students (a random sample of twenty-five, for example) would reveal many different patterns of change. Some students steadily increase, some decrease and then increase, and others increase and then decrease. Such patterns suggest individual pathways to mastery—some are adaptive in terms of helping students arrive at a good outcome, and others appear to be less adaptive, calling for remediation, or even better, early intervention, to stay on a positive curve.

### 4. Assessments that Support Entrustment Decisions

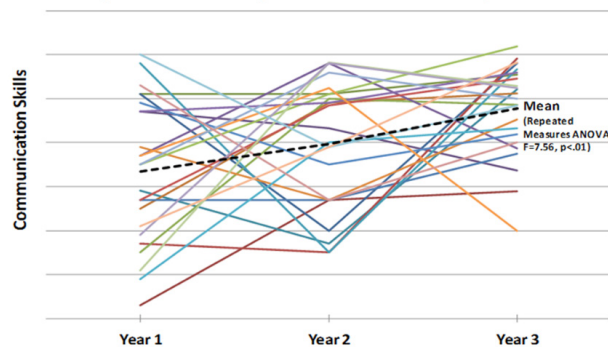
NYU is piloting the implementation of four EPAs as part of the AAMC’s national program for the core EPAs for entering residency: (1) prioritizing a differential diagnosis following a clinical encounter; (2) forming clinical questions and retrieving information to advance patient care; (3) collaborating as a member of an interprofessional team; and (4) identifying system failures and contributing to a culture of safety and improvement. Our focus has been on inventorying the curriculum to ensure sufficient education and training for those specific EPAs and

**Figure 8: Progressive Mastery of Communication Skills**  
Communication Skills Over Time



**Figure 9: Individualized Competency Pathways**

OSCE-Assessed Communication Skills Throughout Medical School  
(n=25, Random Sample of 170 Medical Students)



reviewing existing assessment data to determine whether there is sufficient information for faculty to decide if they can “entrust” students to be able to perform these tasks on the first day of residency.

We have a reasonable amount of assessment information on students’ competencies related to at least two of the EPAs: (1) prioritizing a differential diagnosis following a clinical encounter and (2) collaborating as a member of an interprofessional team. However, we lack sufficient information based on actual clinical encounters and therefore are developing and piloting new workplace-based or point-of-care assessments that will have faculty provide quick structured feedback on multiple patient encounters throughout students’ rotations. These assessments are meant to be “frictionless” in that they do not take more than a few minutes, can be done in the moment, at the bedside, and do not ask faculty to make broad summative assessments—instead, they focus on describing observable behaviors essential to the relevant EPA and making an intuitive interim assessment about the entrustability of the student based solely on the authentic clinical interaction. Technology makes this possible because we use mobile devices to prompt and collect these assessments. Essentially, the power of technology allows the kinds of everyday observations and expert judgments that are ubiquitous in faculty supervision of medical students to be captured and shared with students and educational advisors. This occurs across patient experiences within a clerkship, across clerkships as the student learns about different specialties, and over time to ensure convergence upon entrustability.

We also recognized the need to collect similar kinds of workplace-based data for the EPA on interprofessional collaboration and, perhaps even more importantly, to include a multi-source perspective on teamwork competencies.<sup>77</sup> We are therefore piloting a workplace-based assessment of students’ competence in providing patient care as part of a team by eliciting structured feedback from nursing staff, other relevant members of the care team (such as respiratory therapists, physician assistants, and social workers), faculty, peers, and patients. Collating and displaying these data across curricular experiences and over time is the challenge we plan to tackle next. This involves determining how best to display many formative data points, isolated summative assessments, data from across many sources and from within many curricular contexts, and assessments with varying degrees of quality or uncertainty associated with them. Once the data are compiled and decisions made about how best to display the data, we will begin piloting the process of reviewing the data; identifying students who appear to be in need of additional resources, support, and coaching; providing students with the data and exploring whether access to such rich, formative, and varied assessment data facilitates reflection and deliberate practice; and finally, using these data to decide if students are entrustable to perform the selected professional activity. Figure 10 depicts a conceptual prototype we have developed for displaying all the assessment data related to the EPA of prioritizing a

---

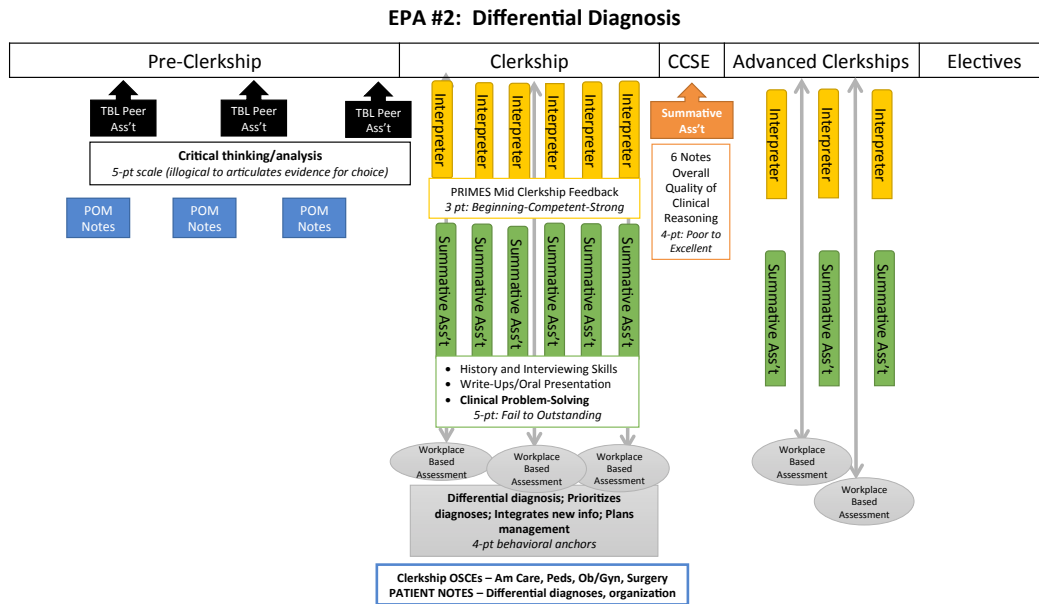
77. These multi-source perspectives are often referred to as 360-degree assessments. Emily E. Anderson & Mark Kuczewski, *The Search for a Meaningful Evaluation of Professionalism*, in HANDBOOK ON MEDICAL STUDENT EVALUATION AND ASSESSMENT 147, 153 (Louis N. Pangaro & William C. McGaghie eds., 2015).



## LEARNING OUTCOMES AND PERFORMANCE IN MEDICAL SCHOOL

differential diagnosis throughout medical school—a view of these data could help students track their progress, educators monitor and support student progress, and researchers understand how entrustability develops and is sustained.

**Figure 10: Prototype for Dashboard with Multiple Data Sources for Determining “Entrustability” to Perform EPA #2—Prioritizing a Differential Diagnosis**



Finally, while the ultimate outcome of medical school training should be assessed through graduates’ influence on the health of their patients, we have not yet convincingly demonstrated links between patient outcomes and educational efforts because of the messiness of our data, the number and complexity of confounding factors, and the many determinants of patient health that are well outside an individual physician’s sphere of influence.<sup>78</sup> While we have not given up on that quest and have identified some patient outcomes that are potentially both measurable and sensitive to educational activities, in the shorter term, we have turned to the “recipients” of our graduates—residency program directors—and asked them to assess the readiness of our students for practice. Annually, we survey the directors of each residency program that has a graduate of NYU School of Medicine and ask them to rate student competencies after a year of residency training. These assessments are particularly informative in terms of identifying strengths and weaknesses in our curriculum as well as individual graduates who are struggling in residency.

78. See Adina L. Kalet et al., *New Measures to Establish the Evidence Base for Medical Education: Identifying Educationally Sensitive Patient Outcomes*, 85 *ACAD. MED.* 844 (2010) (discussing the lack of evidence and need to benchmark patient outcomes to evaluate the effectiveness of medical education).

## V. THE CRITICAL ROLE OF SIMULATION: NYU'S STANDARDIZED PATIENT (SP) PROGRAM

While all the assessment examples mentioned above are important to achieving a “program of assessment,” none are as critical as our SP Program. The standardized or controlled nature of simulation allows us to understand how individual students perform in comparison with their peers and to track performance developmentally, permitting the identification of both developmental trends and individual patterns. The ability to design clinical scenarios and challenges also provides educators with the ability to present students with developmentally appropriate tasks, delivering the right clinical problem at the right level at the right time to reinforce or extend students’ skills, and focusing either on specific elements of a clinical task or the bundled, integration of such elements into an authentic, clinical “whole.”

### A. Using Actors to Train the Next Generation of Physicians

Simulation with trained actors, known as SPs, provides rich educational opportunities, as there are many hard to master clinical skills that our medical students need to learn. Physicians need to integrate all the information about medical history, personal information, the physical exam, and diagnostic testing to determine what is going on and what to do about it. They also need to be professional and compassionate while listening carefully, exercising clinical reasoning, establishing trust, and determining next steps. We all want our learners to practice before they actually have to perform these complex tasks in real life when the stakes are higher.

SPs and OSCEs play a more extensive role than ever within the modern medical education curriculum.<sup>79</sup> SPs and OSCEs are important educational tools for high quality teaching and practice (formative assessments) as well as for the evaluation of basic and advanced clinical skills (summative and formative assessments).<sup>80</sup> They are the building blocks of interactive medical education in medical schools, residency programs, and increasingly, continuing education for faculty.<sup>81</sup> In addition, medical students in the United States must now all pass a certifying exam<sup>82</sup> that includes interaction with twelve SP cases to get their license.<sup>83</sup>

79. See *Step 2 CS, U.S. MED. LICENSING EXAMINATION*, <http://www.usmle.org/step-2-cs/> (last visited Jan. 30, 2018). OSCEs in which SPs—highly trained actors—portray standardized clinical scenarios that call for learners to demonstrate their clinical skills and competence have been used in medical education since the 1970s and are now a required element of the U.S. Medical Licensing Examination for all U.S. medical graduates. *Id.*; see also *OBJECTIVE STRUCTURED CLINICAL EXAMINATIONS: 10 STEPS TO PLANNING AND IMPLEMENTING OSCEs AND OTHER STANDARDIZED PATIENT EXERCISES 2* (Sondra Zabar et al. eds., 2013) [hereinafter *CLINICAL EXAMINATIONS*].

80. *CLINICAL EXAMINATIONS*, *supra* note 79, at 1.

81. *Id.*

82. FED’N OF STATE MED. BDS. OF THE U.S. & NAT’L BD. OF MED. EXAM’RS, *STEP 2 CLINICAL SKILLS (CS) 3* (2017), <http://www.usmle.org/pdfs/step-2-cs/cs-info-manual.pdf>.

83. See *id.* at 4–10.

## LEARNING OUTCOMES AND PERFORMANCE IN MEDICAL SCHOOL

Simulation allows us to create all kinds of complex scenarios that may include specific, targeted skills (such as performing a procedure), focus on one aspect of a clinical encounter (such as gathering a history), or involve a full complement and integration of the variety of clinical skills needed to be an effective physician. Communication skills are considered central to all these types of clinical challenges, and over time, the field has come to recognize that these skills are learnable, requiring practice, guidance, and feedback.<sup>84</sup>

At NYU School of Medicine, we have built an extensive simulation program to use actors to train better doctors. Simulation is integrated into the curriculum on the first day of medical school, continuing through the curriculum in the core “doctoring skills courses” (Practice of Medicine and Physical Diagnosis) and into the clinical clerkship years. NYU School of Medicine started using actors as SPs in 1995, initially for forty students, and now for all 640 of our medical students, in thirteen residency programs, and for onboarding of faculty. There are close to one thousand learners annually participating in simulation. We have worked closely with the medical school and graduate medical education programs to ensure that we are teaching common core skills and evaluating them all in the same way.

For example, in the first week of medical school, the new students meet Rose, a seventy-eight-year-old just discharged from the hospital after having a heart attack. She smokes and the student is challenged to talk to her about quitting. They also see Pedro, an eighteen-year-old with Type 1 diabetes, who removes his insulin pump when he plays basketball because he is embarrassed and has recently ended up in the hospital as a result. Finally, they meet Robert, a vegan freelance writer with a strong family history of colon cancer and an unwillingness to get a colonoscopy because he is uninsured. By starting with these types of patients the first day of medical school, learners are sensitized to the skills they will need to become a competent physician. Throughout medical school, students will encounter more than forty different SPs. We have designed and implemented over 500 cases, designing some to address clinical conditions or patients they are unlikely to experience within the local patient populations, others to reflect the “bread and butter” clinical situations physicians face every day. Cases can be structured to focus on isolated skill sets but more often integrate communication skills, medical knowledge, clinical reasoning, and the need to attend to patients’ psychosocial contexts.

NYU School of Medicine has a 25,000 square foot simulation center, which is a public-private partnership between the medical center and the City University of New York.<sup>85</sup> The physical space includes fourteen rooms in which to perform patient/health professional interactions, thirty-seven mannequins and partial task trainers, and other flexible spaces that can be used to simulate operating rooms, hospital rooms, family meetings, or even large-scale disasters.

---

84. Sherrie H. Kaplan et al., *Assessing the Effects of Physician–Patient Interactions on the Outcomes of Chronic Disease*, 27 *MED. CARE (SUPP.)* S110, S110 (1989); see also Richard L. Street Jr. et al., *How Does Communication Heal? Pathways Linking Clinician–Patient Communication to Health Outcomes*, 74 *PATIENT EDUC. & COUNSELING* 295, 297–99 (2009).

85. *About*, NYSIM, <http://nysimcenter.org/about> (last visited Jan. 30, 2018).

Hiring and training actors for educational simulation is possible in any environment. In urban settings like New York City, it is especially easy.<sup>86</sup> Actors are paid twenty-five dollars an hour and are given extensive training how to portray the patients—including their personal history, personality/temperament, and their likely responses to questions. Equally as important, they are trained to use a behaviorally anchored checklist to reliably assess and give feedback to our learners. SPs typically train for four to six hours per case. Some of our SPs have seen over 100 learners work through the same clinical cases and therefore have become highly calibrated. Understanding the variations in ability among trainees makes SPs expert at providing feedback and making fair, accurate, and reliable assessments. We currently have more than 300 actors who work for the medical school regularly. Based on these experiences, we have published a manual detailing how to design and implement SP exercises.<sup>87</sup>

Each scenario typically lasts ten to fifteen minutes. After the scenario is complete, the SP rates the performance of the learner, and then there is time for feedback either from the SP or from a faculty observer who viewed the interaction from outside the room via video or the one-way observation window. At NYU School of Medicine, we also use simulation with teams and can focus on specific interprofessional activities, which have allowed us to contribute well-controlled, standardized assessments of teamwork skills to our EPA dashboards.

Since all our cases are designed using the same framework and behaviorally anchored, standardized checklist, clinical skills reports are created. These reports all roll up into the learner dashboard. Students are able to see their performance compared to their class's average in three domains: communication skills, information gathering, and patient education. Individual learning plans can be created for each learner and then reviewed and compared to their performance over another series of SP cases.

These health professional simulations highlight the patient's perspective and experience and remind learners that the ultimate outcome of education is a healthy, activated, satisfied patient. It is rare for physicians or any health professionals to get specific feedback from their patients because patients tend to fall into several categories that make it hard to assess variation in patient satisfaction: patients who love their physicians or feel indebted to them, patients who do not want to be negative and try to give their physicians the benefit of the doubt, patients who are so unhappy with their physician that they immediately find a new one (and therefore often are not included in patient surveys), and patients who do not know what standards to hold their physicians to and therefore are generally satisfied with what many would regard as sub-standard care. SPs have the breadth of experience and the deep understanding of the behavioral standards expected of physicians (as clearly laid out in the assessment tools we use) to provide very detailed, specific, and constructive feedback to learners. For example: "The student seemed to be very nervous and spoke

86. Elizabeth Zimmer, *Playing Sick: How Actors Are Making Better Doctors*, VILLAGE VOICE (Aug. 31, 2016), <http://www.villagevoice.com/arts/playing-sick-how-actors-are-making-better-doctors-9042319/>.

87. See generally CLINICAL EXAMINATIONS, *supra* note 79 (offering the systematic approach to make it easier for more people to get involved in the process of creating OSCEs).

really fast.” “Student was not organized and was so tense it made it hard for the patient to connect with him and form a trusting relationship.” “The student was a very effective communicator and was very good at developing a rapport with the patient—however, she completely missed the detail about my use of the medication or did not understand the implications and so was never really able to get to the issue and help the patient figure out what to do next.”

This type of genuine, straightforward feedback, rarely available in actual practice, can motivate learners to change. Our data suggest that having these repeated opportunities to practice and get feedback on communication skills leads to improvements in communication skills throughout medical school.

Understanding how SPs’ ratings correlate with learners’ impact on real patients is a critically important education research question. If learners can motivate an actor in a simulation, can they motivate real patients to lose weight? We conducted a study to answer just this question and found that resident physicians who had been found to be effective weight loss counselors with SPs in simulated cases had “real” obese patients who lost more weight than the obese patients of residents who had not been effective counselors with SPs.<sup>88</sup>

In 2009, we implemented an “Unannounced” SP program in which we send “secret shoppers”<sup>89</sup> into our clinical environments. This has provided a rich and unparalleled source of data on how our clinical health system functions and our learners perform within that system. Based on these encounters, we have provided routine feedback reports to the healthcare system and to our educators and learners. This project also allows us to compare performance in “announced” clinical exams with SPs to performance in actual clinical settings when the resident physician is dealing with the everyday chaos and complexity of the healthcare system. This provides insight into the degree to which skills assessed in simulation transfer to the real world, suggesting that our actors can predict what is likely happening in authentic settings, behind closed clinic doors.

In medical education, actors have become critical to physician training. They help the learners develop lifelong learning and deliberate practice skills and enable educators to identify struggling learners and create individual learning plans. Performance as assessed by SPs appears to reflect how health professionals practice in real world settings. And we are currently engaging in additional research to identify the facilitators and barriers to the transfer of skills into real practice environments. Ultimately, SP activities are central to the development and assessment of students’ clinical skills and provide feedback on students’ progressive mastery of those skills.

---

88. Melanie R. Jay et al., *The Impact of Primary Care Resident Physician Training on Patient Weight Loss at 12 Months*, 21 *OBESITY* 45 (2013).

89. Secret shoppers are SPs whom the physician and healthcare team believe to be real patients. See Sondra Zabar et al., *Unannounced Standardized Patients: A Promising Method of Assessing Patient-Centered Care in Your Health Care System*, 14 *BMC HEALTH SERVS. RES.* 157 (2014).

## VI. CONCLUSION

Medical and legal education professionals agree on the importance of experiential learning and skills development and the role of assessment and feedback in fostering learning and skills mastery. The professions have much in common. In this paper, we share general assessment trends in medical education and our experience in creating a comprehensive program of assessment in one medical school. We have found that thinking of assessment as part of a broader, rationally organized, and consistently delivered “program of assessment” helps clarify the goals of assessment and the meanings and interpretations attached to assessment data. It also reinforces the importance of supporting students in reflection and the development of expertise. All this provides students with the curricular content and activities necessary to learn and become competent, and ultimately, to be entrusted to perform the core duties of the profession as expected by our social contract.