Sondra Zabar · Elizabeth Kachur
Kathleen Hanley · Colleen Gillespie
Adina Kalet *Editors*

Objective Structured Clinical Examinations

10 Steps to Planning and Implementing OSCEs and Other Standardized Patient Exercises

Second Edition

Springer

Sondra Zabar • Elizabeth Kachur
Kathleen Hanley • Colleen Gillespie
Adina Kalet
Editors

Objective Structured Clinical Examinations

10 Steps to Planning and Implementing OSCEs and Other Standardized Patient Exercises

Second Edition

Springer

Editors
Sondra Zabar, MD
Department of Medicine, Division of General Internal Medicine and Clinical Innovation
New York University Grossman School of Medicine
New York, NY, USA

Kathleen Hanley, MD
Department of Medicine, Division of General Internal Medicine and Clinical Innovation
New York University Grossman School of Medicine
New York, NY, USA

Adina Kalet, MD, MPH
Department of Medicine
Medical College of Wisconsin
Milwaukee, WI, USA

Elizabeth Kachur, PhD
Medical Education Development
Global Consulting
New York, NY, USA

Colleen Gillespie, PhD
Institute for Innovations in Medical Education, Department of Medicine
New York University Grossman School of Medicine
New York, NY, USA

ISBN 978-3-031-88751-2 ISBN 978-3-031-88752-9 (eBook)
https://doi.org/10.1007/978-3-031-88752-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

I introduced the OSCE 50 years ago in response to the need for a tool that assessed, more objectively than had been previously possible, the range of clinical skills expected of a student. Since then, the OSCE has been established as the gold standard for performance with recognition of its validity, measuring what we should measure; its reliability with consistency and dependability; its practicality, both flexible and adaptable; its impact on learning, influencing students and teaching; and its acceptability, being well received by students and teachers.

The first edition of *Objective Structured Clinical Examinations: 10 Steps to Planning and Implementing OSCEs and Other Standardized Patient Exercises* made an important contribution to the health professions education literature. It provided the different stakeholders with an understanding of the principles underpinning an OSCE and how these are reflected in practice. It was well received. It demonstrated for experienced examiners how they could gain the maximum benefit from the approach. For newer staff and students, it provided a sound and informative introduction to the use of the OSCE as an assessment tool.

Health professions education does not stand still and has continued to evolve since the first edition was published. The COVID-19 pandemic has had a profound effect on teaching and learning and on the assessment of students, with a move to online and hybrid approaches combining elements of face-to-face and online experiences. There has been a reappraisal of the learning outcomes to be assessed and of the use of technology in assessment.

The new edition recognizes these challenges and the implications for conducting an OSCE in different countries around the world. The OSCE, with its multiple samples of performance, continues to dominate performance assessment as a key approach to formally assessing the clinical skills of undergraduate and postgraduate students. The book continues to serve as a valuable resource to support best practice in assessment and in the use of the OSCE. It highlights a greater virtual element in assessment, the expansion of formats to address new areas such as telemedicine, the need for greater student contributions and approaches to remediating poorly performing students, and the move from assessment *of* learning to assessment *for* learning, particularly relating to the various transitions facing a student. The book recognizes the need for scholarship and research in education and how these are exemplified in the OSCE. A range of practical approaches, including the use of simulated patients in real clinical environments, is explored in this second edition.

The assessment of a learner's clinical competence is almost certainly the most important responsibility for the teacher or trainer. This new edition can help prepare all concerned with the necessary skills. As in the first edition, the text has been carefully crafted and will continue to be of value to the more experienced and novice examiner.

Professor of Medical Education (Emeritus)
University of Dundee
Dundee, UK

Ronald M. Harden
PhD,OBE,MD,FRCP,FRCS(Ed),FRCPC

Preface

The OSCE Book second edition is a practical manual to guide and support health professions educators wishing to build or improve a state-of-the-art performance assessment program for trainees or faculty.

This edition builds on the first in several key ways. First, it introduces new methods—virtual Objective Structured Clinical Examinations (OSCEs) and group OSCEs, as well as new applications—OSCEs for teaching, health professions transitions, and telemedicine. Second, it highlights the scholarship underlying, supporting, and implementing the methods, in particular, the use of notes to assess clinical reasoning and the variety of pragmatic performance-based assessment "use-cases" across health professions education. Finally, we attempt to provide enhanced sophistication to organizing, visualizing, and reporting the data from these experiences to optimize learning for individuals, programs, and institutions. We aim to provide a very pragmatic book that offers hard-earned, expert advice and concrete material, which can be used as is or tailored to local needs.

Since the first edition in 2013, myriad medical schools have chosen to undertake the serious investment of effort, time, and resources needed to do performance assessment well. They have chosen competencies to measure, written and piloted cases, recruited and trained standardized patients/participants (SPs), developed standards, and scheduled hordes of students into countless 15-min rotation slots many times a year. Why? Because of the growing acceptance that clinical knowledge, which can be reliably and validly tested through multiple-choice exams, does not translate directly into clinical skill. Because after assimilating new knowledge, the health professions student must learn to integrate and apply their cognitive learnings in human encounters together with communication, professionalism, ethical behavior, clinical reasoning, and physical examination skills. To measure these complex, higher-order behaviors, Objective Structured Clinical Examinations (OSCEs) are needed to simulate—physically and emotionally—actual health professional–patient encounters.

This book emerged from a handout for a workshop at the Association for Program Directors in Internal Medicine annual meeting on developing SP programs. Attendees suggested useful expansions and encouraged publication. The first edition was a comprehensive practical guide to developing SP exercises in physician training. It explained how to design OSCE cases, how to recruit and train SPs, manage logistics, and do all the nitty-gritty things that make an SP program sing or stumble. At the time, this book filled a gap so wide globally that it was translated into Chinese.

The editors and authors of this book are our partners who build and lead the NYU/ Bellevue Primary Care Residency Program. They are well positioned to create this manual because of their extensive experience, dedication, and pioneering scholarship in medical education. Starting small in the late 1980s, we experimented with SP encounters in our doctoring course for medical students and in the Primary Care Internal Medicine Residency Program. By 2000, we had begun to use SPs for formative and summative educational experiences on a large scale and across a broad range of training levels and content areas including geriatrics, women's and immigrant health, and addiction medicine. We gained experience in creating research-quality OSCEs to assess communication skills training in the multi-institutional Macy Initiative in Health Communication project (medical students) and then in a disaster preparedness project

on psychosocial aspects of bioterrorism jointly funded by the Centers for Disease Control and the Association for American Medical Colleges (physicians, nurse practitioners, and physician assistants). We developed complimentary baseline, end-of-clerkship-year, and near-graduation medical student clinical encounter skills assessments to allow us to understand how clinical skills progress in novice healthcare providers. To do this, we committed to using a single measurement framework for every assessment regardless of stage; and we committed to monitoring progress longitudinally from Undergraduate to Graduate Medical Education. In the mid-2000s we began to pilot the use of unannounced SPs in our residency program in order to understand how what we measure in OSCEs translates to the real practice setting, enabling us to follow our trainees from the simulated into the "real" clinical setting. Currently, our medical students encounter upwards of 40 SP cases during medical school and our primary care residents complete around 45 cases in 3 years of training. These encounters range from formative exercises designed purely for learning, which include immediate feedback and extensive debriefing, to summative, high-stakes exams—increasingly, in the actual clinical setting. Each case or set of cases tests multiple dimensions (e.g., preparation, communication, clinical reasoning, time management, preventive medicine, error prevention, and management) of good doctoring and approaches the complexity of caring for real patients and the stresses of actual clinical practice in a controlled, reproducible, and measurable setting.

After some initial resistance to the idea of OSCEs, our students and residents early on expressed their appreciation for the opportunity to practice difficult tasks in a safe environment. Now, although students still get nervous in anticipation, our OSCEs are popular and perceived as valuable teaching tools. Our residents, by and large, rate them highly, feeling they are an efficient use of time and excellent learning experiences. Faculty who participate in OSCEs—developing cases, observing, giving feedback and debriefing—report benefitting greatly from the opportunity to directly observe and calibrate their expectations and understanding of trainees. Clinical leadership appreciates that we are rigorously addressing important issues such as communication, patient safety, and patient activation. And the many SPs working closely with us in this endeavor feel that they are engaged in meaningful work, both personally and professionally. In the Primary Care Internal Medicine Residency Program, our annual all-program OSCE, initiated in 2000, is not only a central feature of program evaluation and resident assessment, but has also become an important community-building experience for residents, staff, and faculty.

OSCEs are now used in the training of many health professionals in the United States and elsewhere. They are used to assess knowledge, skills, professionalism, ethical behavior, physical examination skills, and the ability to work with difficult patients, patients with diverse cultural backgrounds, patients on the phone conducting telehealth, and with families. They can measure simple processes (does the learner recommend stopping smoking?) and very complex ones (does the novice have the professional maturity to manage telling the non-English-speaking family member through an interpreter about an unexpected death and ask for an autopsy?).

Much progress has been made in our ability to ensure that we graduate physicians capable of practicing medicine in our rapidly evolving healthcare environment. Yet, so much more awaits. New curriculum needs are emerging every year—interprofessional education, patient safety, systems-based practice, informatics, disaster medicine, telemedicine, to name a few recent additions. In the way that a crisis can predictably provide many breakthrough opportunities for innovation, the COVID-19 pandemic, which disrupted medical education and clinical practice, accelerated the need for remote or virtual simulation when medical students were temporarily barred from clinical settings because of limited personal protective equipment. Within days, we were able to conduct comprehensive simulations to confirm that our near graduates were ready for internship by having them demonstrate their capabilities with SPs who engaged with them and assessed from the safety of their own homes. In our case, the SPs were trained and calibrated already, and they easily mastered the technology. Students in a rural Midwest state were assessed by SPs in a southern state. These encounters were directly

observed and debriefed in groups on a variety of virtual software platforms. In the end, students and faculty thanked us for continuing their education under dire circumstances.

Having a rich and flexible SP-based OSCE program has allowed us to meet new curriculum and assessment challenges in a rigorous and exciting way. Developing and implementing an OSCE is a highly creative and scholarly activity, which requires a group of educators to do the difficult work of coming to consensus on educational priorities and setting standards for trainee performance. The process is scholarly because, when engaging in OSCE development, one confronts important unanswered questions about health professional competence and training. The OSCE is a robust method for answering many of these questions.

For all these reasons, we find this work enjoyable and intellectually engaging, and we believe you will too. In this book, the authors share with you their experience, which will allow you to avoid common pitfalls and get directly to the fun and meaningful stuff. Call us, come visit, come see OSCEs in action, and organize a workshop. We stand ready to help.

New York, NY, USA — Mack Lipkin, MD

Milwaukee, WI, USA — Adina Kalet, MD

New York, NY, USA — Sondra Zabar, MD

Acknowledgments

First of all, we must thank the thousands of NYU Grossman School of Medicine medical students, residents, chief residents, fellows, and faculty—from Primary Care, Categorical Medicine, Surgery, Orthopedics, Anesthesiology, OB/GYN, Pediatrics, and more—who participated in OSCEs over the past decades. We appreciate their trust that the SP exercises we introduced would be educational and relevant. We are immensely grateful to the many creative and dedicated colleagues on the NYUGSOM faculty, in addition to this book's contributors, we are indebted to the early adopters who enthusiastically wrote cases, observed and evaluated learners, trained countless SPs, and conducted faculty development, especially Jennifer Adams, Richard Greene, Anne Dembitzer, David Stevens, Margaret Horlick, Barbara Porter, Patrick Cocks, Kevin Hauk, Abigail Winkel, and Arielle Kurzweil. thanks are due to Ellen Pearlman and Judith Brenner, who designed the workshop that included the first incarnation of this book as a manual more than 15 years ago! We are also indebted to Michael Yedidia for the initial formulation of our long standing OSCE communication skills checklist, now dubbed the Clinical Communication Skills Assessment Tool (CCSAT). Much appreciation to Ronald Harden for writing the Foreword and Mack Lipkin for writing the Preface. Over the years, a few key administrative and organizational geniuses were essential to the success of each OSCE. We would like to thank the inimitable Marian Anderson, Amy Hsieh, Regina Richter, Julianne Cameron, Lindsey Disney, Ivey Overstreet, Katherine Miller, Sienna Kurland, Davis Boardman, and our SP Educators Ari Kreith, Virgina Drda, and Meg Anderson, who have cast and directed hundreds of SPs for our many programs. You all deserve an Academy Award.

This second addition, appearing 10 years after the first, truly could not have been accomplished without the dedication of interdisciplinary faculty members and research staff of our Program for Medical Education Innovations and Research—including Naya Gonzalez, Jeffrey Wilhite, Christine Beltran, and countless summer scholars and students. Our data scientist colleagues—Tavinder K. Ark, Ilan Reinstein, Nikola Koscica, Nivedha Satyamoorthi, and Khemraj Hardowar—have made many of our initial visions of a robust case bank and longitudinal, performance-based assessment database come to life well beyond our imagination. Together, this group raised our ability to set standards for educational scholarship on performance-based assessment, created reports and dashboards that ensured that our students and trainees received valuable, actionable feedback, and disseminated abstracts, workshops, and materials that helped us share our work with the larger medical education community. They were the impetus for all the updates in this book.

We are deeply grateful to Harriet Fisher, Zoe Phillips, and Renee Heller for keeping us on track for this second edition—their intellect, careful copy editing, organizational skills, and can-do attitudes were essential to our success and priceless.

We are incredibly fortunate to work at institutions who value the importance of investing in medical education, communication skills, assessment, and scholarship to improve patient outcomes, including the New York Simulation for Health Sciences (NYSIM), Gouverneur Health,

Bellevue Hospital Center, and NYU Grossman School of Medicine and its Office of Medical Education, all of whom supported this project with material and collegial support from day one. Finally, we would like to thank the many standardized participants who have partnered with us to ensure that medical education is accountable to our ultimate goal of serving our patients, communities, and society with excellence and compassion.

Contents

Contributors

Jennifer Adams, MD Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Kathleen Allen, MPA New York Simulation Center for Health Sciences, New York, NY, USA

Lisa Altshuler, PhD Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Tavinder K. Ark, PhD, MPH Data Science Institute, Medical College of Wisconsin, Milwaukee, WI, USA

Christine Beltran Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, USA

Brielle Blatt, MPA New York Simulation Center for Health Sciences, New York, NY, USA

Lynn Buckvar-Keltz, MD Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Calvin L. Chou, MD Department of Medicine, University of California San Francisco School of Medicine, San Francisco, CA, USA

Ruth Crowe, MD, PhD Department of Medicine, New York University Grossman Long Island School of Medicine, New York, NY, USA

Anne Dembitzer, MD Department of Medicine, Division of General Internal Medicine and Clinical Innovations, New York University Grossman School of Medicine, New York, NY, USA

Virginia Drda New York Simulation Center for Health Sciences, New York, NY, USA

Kinga L. Eliasz, PhD Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Rachel H. Ellaway, PhD Department of Community Health Sciences, University of Calgary Cumming School of Medicine, Calgary, AB, Canada

Harriet Fisher, MD Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Colleen Gillespie, PhD Institute for Innovations in Medical Education, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA

Richard Greene, MD, MHPE Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Kathleen Hanley, MD Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Katherine Hochman, MD Division of Hospital Medicine, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA

Renee Heller Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Elizabeth Kachur, PhD Medical Education Development, Global Consulting, New York, NY, USA

Adina Kalet, MD, MPH Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA

Arielle Kurzweil, MD Department of Neurology, New York University Grossman School of Medicine, New York, NY, USA

Katharine Lawrence, MD, MPH Department of Population Health, New York University Grossman School of Medicine, New York, NY, USA

Willie Lombeida New York Simulation Center for Health Sciences, New York, NY, USA

Gene Mamaril New York Simulation Center for Health Sciences, New York, NY, USA

Jeffrey Manko, MD Department of Emergency Medicine, New York University Grossman School of Medicine, New York, NY, USA

Cynthia Osman, MD Department of Pediatrics, New York University Grossman School of Medicine, New York, NY, USA

Sean Overstreet New York Simulation Center for Health Sciences, New York, NY, USA

Donna Phillips, MD New York Simulation Center for Health Sciences, New York University Grossman School of Medicine, New York, NY, USA

Zoe Phillips, MD Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Barbara Porter, MD, MPH Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Daniel J. Sartori, MD Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA

Verity Schaye, MD, MHPE Institute of Innovation in Medical Education, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA

Katherine Tame New York Simulation Center for Health Sciences, New York, NY, USA

Linda Tewksbury, MD Department of Pediatrics, New York University Grossman School of Medicine, New York, NY, USA

Andrew Wallach, MD Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Elizabeth Wargo Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Elizabeth Weinshel, MD Division of Gastroenterology, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA

Jeffrey Wilhite, MPH Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Renee Williams, MD, MHPE Institute for Excellence in Health Equity, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA

Sondra Zabar, MD Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

Introduction: OSCEs: Past, Present, and Future

1

Sondra Zabar, Elizabeth Kachur, Kathleen Hanley, Colleen Gillespie, and Adina Kalet

Creating Objective Structured Clinical Exams (OSCEs) or other Standardized Patient/Participant (SP) exercises can feel overwhelming, but the benefits of this type of instructional design regarding practice-based learning and performance-based assessment for future healthcare practitioners certainly make OSCEs worthwhile. This is why we wrote this book. It is our hope that the systematic approach offered here will make it easier for more people to get involved in the process of creating OSCEs or similar SP exercises. Using Chap. 2, Organizing OSCEs (and Other SP Exercises) in Ten Steps, as a road map, the process is doable and rewarding. Furthermore, the publication of the first edition of this book in Mandarin by Peking University Medical Press (2018) illustrates the worldwide need for disseminating basic and more advanced strategies for implementing such programs.

This vastly expanded second edition is a testament to how much the medical education field around OSCEs has grown. Organizations such as the Association for SP Educators (ASPE) and the Society for Simulation in Healthcare (SSH) have consistently fostered the development of research and the creation of standards. A simple PubMed search using the term "OSCE" resulted in a total of 2791 references (queried on July 26, 2023). Figure 1.1 and Ba et al. [1] illustrate this rapid growth. Over the course of 10 years (2012 when the first edition was completed to 2022), the number of articles indexed per year more than doubled. This is surely a low estimate, since not all journals are indexed in PubMed, and "OSCE" is not always the term used to describe such multi-station clinical exams or exercises.

S. Zabar (✉) · K. Hanley
Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA
e-mail: sondra.zabar@nyulangone.org

E. Kachur
Medical Education Development, Global Consulting, New York, NY, USA

C. Gillespie
Institute for Innovations in Medical Education, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA

A. Kalet
Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA

Definitions

Objective Structured Clinical Exams or Exercises (OSCEs)

OSCEs are training or assessment programs in which learners rotate through a series of time-limited stations. At each station, a learner has to perform specific tasks that are kept constant across all trainees. Rating forms with predetermined performance criteria are then used to assess the learner's skills in a standardized fashion. Figure 1.2 illustrates the SP cases a learner might encounter in an eight-station OSCE (further expanded as the example OSCE in the blueprint in Table 2.6).

Over the years, many additional terms have surfaced such as Clinical Practice Exam (CPX), Comprehensive Clinical Exam (CCE), Clinical Skills Assessment (CSA), or Objective Structured Clinical Assessment (OSCA). To distinguish themselves from other institutions, educators developed local terminology to label their Multiple Station Exam/Exercise (MSE). More recently, there have been some questions about whether it is indeed possible to make OSCEs truly "objective," given the human tendency toward subjectivity. Thus, some educators are advocating for replacing "objective" with "observed," a less controversial designation.

Additional terms have developed for specific types of OSCEs. For example, GOSCE (Group OSCE) refers to multi-station exercises where learners rotate in groups. OSTEs are OSCEs that focus on "teaching"; they are used as faculty development or program evaluation tools for continuing professional training (for more information, see Chap. 11).

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_1

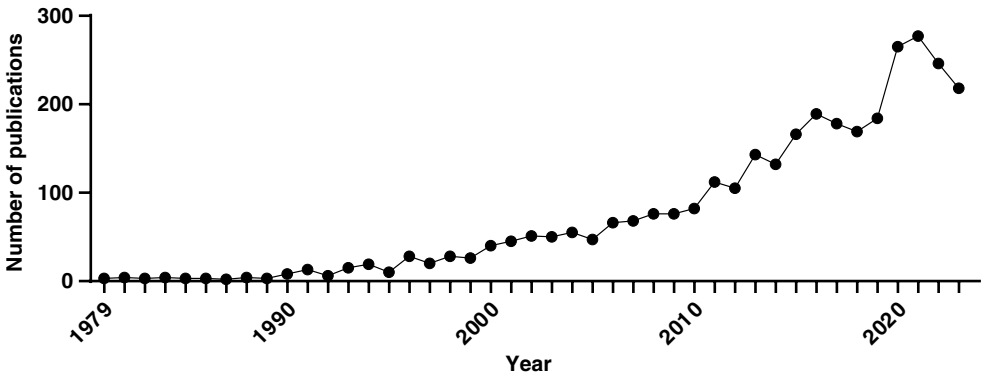


Fig. 1.1 Growth of OSCE-related articles indexed in PubMed, 1979–July 2023. Search query = OSCE

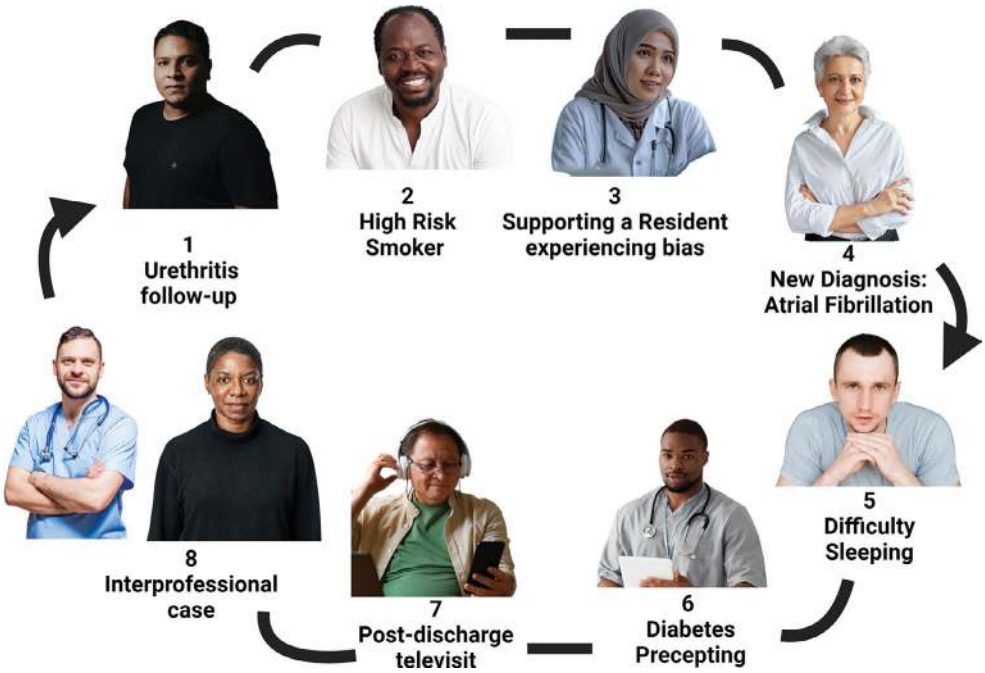


Fig. 1.2 Eight-station OSCE: Circuit of SP scenarios (i.e., stations) through which learners rotate. This OSCE is used as an example throughout Chap. 2, Table 2.6 describes each station. The images accompanying each station depict the standardized participant (patient, trainee, nurse, etc.) needed to portray the case

OSCE terminology can create some confusion for reviewers of the literature. In some papers, "OSCE" is used for single-station programs. This defies the underlying assumption that multiple sampling of performance is needed to create a reliable assessment tool. Thus, by definition, an OSCE is a multi-station event and will require at least two stations. There have been debates about how many stations are needed to arrive at defensible promotion decisions; formative assessments can manage with fewer stations. In fact, too many stations can create cognitive overload which could be counterproductive. A few stations with thorough debriefing can be a more powerful training opportunity than a myriad of stations that exhaust rather than enhance competencies.

Standardized Patients/Participants (SPs)

When SPs were introduced to medical education by Howard Barrows in 1963, they were called "programmed" patients [2] to reflect the educator's ability to shape the scenarios in order to meet curriculum or assessment needs. In the 1980s, the term "simulated patient" became popular. With increasing use in assessment and the corresponding need for controlling the test stimulus, "standardized patient" is often the preferred term, especially in North America. In other continents, the term "simulated patient" remains most prevalent.

However, this terminology is again undergoing some official changes. OSCEs and other simulation-based programs have long surpassed the sole inclusion of patient-based encounters, resulting in the most recent term, "standardized participant." In addition to family members, programs utilize standardized healthcare professionals and their students, as well as many other individuals that may be critical to learning and assessing professional competencies. Typically, SPs are not afflicted by whatever conditions or problems they are depicting. Rather, they are simulating clinical problems, healthcare provider conflicts, or learner issues for the purpose of training and assessment.

Based on specific SP deployments, terms such as Unannounced Standardized Patients (USPs), Announced Standardized Patients (ASPs), Standardized Learners (SLs), and Standardized Healthcare Providers (SHPs) have also been coined. Chapter 4, "Standardized Patients in the Clinical Setting (USP, ASP)," provides some good examples of how USPs and ASPs can be utilized in training and research, and Chap. 9, "Objective Structured Teaching Exercises (OSTEs) from the Teacher's Perspective: What, Why, When, and How?," illustrates the use of SLs.

The History

OSCEs originated in Dundee, Scotland, in the early 1970s. Ronald Harden (author of our Foreword) and his colleagues published the first article describing these multiple station exams [3]. Commonwealth connections and United Nations grants fostered the initial dissemination around the globe in the late 1970 and early 1980s. Hence, OSCEs arrived in the United States via Canada. It was Ian Hart from Ottawa who introduced OSCEs at an Association of American Medical Colleges (AAMC) conference. In 1985, Hart convened the first "Ottawa Conference" which assembled OSCE and assessment novices and experts to advance the field. Since then, this meeting has taken place biennially all over the world and continues to maintain its original name. To better understand the social history of OSCEs, readers can consult Brian Hodges' book, *Objective Structured Clinical Examination: A Socio-History (Second Edition)*, which explores how discourses of performance, psychometrics, and production have propelled the development of this educational method [4].

In September 1983, Emil Petrusa and his colleagues at the University of Texas Medical Branch (UTMB) in Galveston, TX, mounted the first OSCE in the United States for some 140 internal medicine clerkship students. It consisted of 17 station pairs for a total of 34 stations, each 4 minutes in length. The project was presented at the annual AAMC meeting in the fall of 1984 [5]. Two years later, in the spring of 1986, one of this book's co-editors (Kachur, then at Albert Einstein College of Medicine) organized the first OSCE in the New York City area. Other early adopters in the United States included Southern Illinois University (SIU) and the University of Massachusetts (UMass).

In the 1990s, The Macy Foundation funded a national consortium of 6 regional consortia that comprised a total of 28 US medical schools, in an effort to promote performance-based testing [6]. Macy also funded a three institution communication skills curriculum that was evaluated with OSCEs [7]. All advanced the field in areas such as case and rating form development, scoring practices, exam impact on the curriculum, SP performance quality control, and SP versus faculty observer ratings.

Also in the 1990s, the Educational Commission for Foreign Medical Graduates (ECFMG) developed a growing interest in performance-based assessment to assure adequate clinical competence and English proficiency of international medical graduates (IMGs). This led to extensive pilot testing that further expanded the field (e.g., [8]). By 1998, the ECFMG had created a secure assessment center in Philadelphia, PA, and fully implemented its Clinical Skills Assessment (CSA) as a requirement for all IMGs who wanted to take up postgraduate training in the United States. Similar initiatives were developed in other countries even earlier. For example, in 1975 the United Kingdom General Medical Council (GMC) had already introduced the Professional and Linguistic Assessment Board (PLAB) to evaluate foreign medical graduates who planned to continue training in the United Kingdom as well.

In 2004, 6 years after the implementation of the CSA, the National Board of Medical Examiners (NBME) followed suit and opened five testing centers around the country to incorporate OSCEs in the US Medical Licensing Exam (USMLE). Until the COVID-19 pandemic shutdown, all US medical graduates and all IMGs were mandated to complete the USMLE Step 2 Clinical Skills (Step 2 CS) exam. Between 2004 and 2020, the Board of Osteopathic Medical Examiners (NBOME) administered a similar assessment, the Comprehensive Osteopathic Medical Licensing Examination Level 2—Performance Evaluation (COMLEX-USA Level 2-PE). Initially, both exams were halted because of the pandemic, but then the decision was made to discontinue them altogether. The Medical Council of Canada Qualifying Examination Part II (MCCQE Part II) that was started in 1992 was also stopped with the onset of COVID-19. The implications of these major changes resulted in much debate (see Baker [9], Katsufrakis and Chaudhry [10], and Touchie and Pugh [11], among others). Other countries such as Australia and the United Kingdom continued or expanded the practice of a national performance-based licensing exam. For example, in 2024, the PLAB test will be renamed the Clinical and Professional Skills Assessment (CPSA) and will be a requirement for all UK graduates.

The history of SPs actually started a decade before the advent of OSCEs. In 1963, Howard Barrows, then at the University of Southern California in Los Angeles, hired a healthy woman to simulate the case of a paraplegic patient with multiple sclerosis for his neurology clerkship students. This was the introduction of SPs into medical education [2]. Beginning in the early 1970s, Paula Stillman, then at the University of Arizona, used simulated mothers for teaching interviewing skills. She also created the Arizona Clinical Interview Rating Scale (ACIR) [12], which is still used in some OSCEs today. Barrows and Stillman can be considered the originators of a worldwide movement to use SPs in health professions education.

In 1992, the AAMC organized a national consensus conference on SPs [13]. The Association of Standardized Patient Educators (ASPE) was formed in 2001, creating an international network of professionals devoted to SP work and research. Annual conferences, an active listserv, and an extensive website (www.aspeducators.org) offer the opportunity to exchange resources (e.g., cases, SP contact information, references, moulage techniques to simulate physical symptoms) and to develop Standards of Best Practice guidelines [14], which have been translated into many languages and are currently undergoing a major update with the help of a Delphi study. Simultaneously, ASPE also created best practice recommendations for Gynecology Teaching Associates (GTAs) and Male Urogenital Teaching Associates (MUTAs). These are trained individuals (often working in pairs) who teach or assess sensitive physical exams by using their own bodies. They are particularly useful in OSCE stations that feature content such as breast exams or prostate exams.

Current Use

SPs and OSCEs play an increasingly important role within contemporary education across all disciplines and across the continuum of training, from admission decisions (often referred to as Multiple Mini-Interviews (MMIs; [15]) to promotion decisions. Accrediting bodies have embraced them as important tools for training and assessment (e.g., Holmboe and Iobst [16]). The implementation of performance-based licensing exams spurred major growth in OSCEs and simulation centers at medical schools in order to prepare their trainees for this type of assessment. Now that an infrastructure has been built, some licensing boards are trying to pass the responsibility for readiness-to-practice tests back to the individual institutions. Ensuring standardization in such diverse settings is still under debate.

As Table 1.1 illustrates, how each individual OSCE station can address multiple content areas. Mapping them against the ACGME Core Competencies (2001) [17] and/or the Royal College of Physicians and Surgeons of Canada CanMEDs (2015 Update) [18], both internationally used frameworks, can ascertain how targeted or comprehensive an OSCE will be.

Over time, OSCEs have grown to address many topics, including complex communication, physical exam, and procedural skills such as cultural competence [19–21], genetics [22], gastroenterology [23], substance abuse [24], teaching skills [25], transgender care [26], explicit bias [27], and more. Chapter 8, provides a 20-year history of OSCEs at New York University Grossman School of Medicine. It illustrates the richness of OSCE contributions throughout the continuum of medical education, the potential for scholar-

Table 1.1 Potential coverage of international competency standards through sample OSCE stations

Scenarios	Accreditation Council for Graduate Medical Education standards (ACGME, 2001)					
	Patient care	Interpersonal and communication skills	Professionalism	Medical knowledge	Practice-based learning and improvement	Systems-based practice
	Royal College of Physicians and Surgeons of Canada standards (CanMEDS, 2015)					
	Medical expert, health advocate	Communicator	Professional	Scholar	Leader	Collaborator
Initial work-up of patient with undifferentiated problem (e.g., fatigue, cough)	X	X	X	X		X
Prevention counseling (e.g., smoking cessation, immunization)	X	X	X	X		X
Discuss management of chronic disease with patient	X	X	X	X		X
Telehealth follow-up of lab results (e.g., cholesterol test, PPD)	X	X	X	X		X
Chart review (e.g., discuss chart note indicating medical error with colleague)		X	X	X	X	X
Precept a medical trainee (e.g., physical diagnosis, patient management)	X	X	X	X	X	
Perform an online literature search, and discuss findings with a patient	X	X		X	X	X

Adapted from Kachur (2007) [31].

ship opportunities, and the capacity for enormous creativity. Similarly, Harden et al. [28] used case studies to illustrate the proliferation of OSCE applications worldwide. Chapter 10, illustrates how immersive simulation can be used to prepare learners as they transition from undergraduate medical education to graduate education and enter new health systems. Chapter 12, can help build or expand the capacity of your own SP program.

Future Developments

Worldwide, there are efforts underway to transform time-based education (i.e., requiring a certain length of training in terms of months or years) into competency-based education (i.e., requiring the demonstration of specific competencies as requirement for promotion). Since OSCEs are well suited to address many competencies, they are likely to become an even more prominent teaching and assessment tools in the future.

Initially, OSCEs were all in person, with the exception of a few "telephone stations." However, COVID-19 proved that it was also possible to run assessments online. While this format has some limitations, especially regarding physical exams, it has become quite clear that virtual OSCEs can also offer excellent opportunities to assess and teach clinical skills. Considering the embrace of telehealth, it will be critical for learners to gain and demonstrate proficiency in the virtual clinical modality. Chapter 9, provides more details on how to utilize OSCEs for such types of training. Considering that the coexistence of in-person and online encounters also provides organizational flexibilities (e.g., recruitment of SPs of racial or ethnic backgrounds that are not well represented in one's own community), hybrid-type OSCEs are likely to be the future.

We can predict that further technological advances will also influence OSCE work. Artificial intelligence will be able to help us write and validate OSCE stations, create evaluation forms, analyze videotape performance and post-encounter chart notes, produce teaching points and resources to strengthen OSCE-based teaching, and assist with record keeping and administrative tasks [29]. Advancements in high-fidelity simulators, task trainers, and virtual patients will further advance the field and provide hybrid options such as a combination of SP and mannequin.

The increase in interprofessional education is likely to produce more team-related OSCE stations whether they include learners from other professions or standardized participants. Concerns about professionalism and cultural competence will expand such content areas and create stations or entire OSCEs on related topics.

Intra- and interinstitutional efforts will stimulate more research and exponentially enhance our understanding of this assessment and teaching tool. Professional meetings like the Ottawa Conference, ASPE, and the International Meeting for Simulation in Healthcare will continue to provide a forum to exchange ideas and enhance OSCE methodologies.

How to Use This Book

Chapter 2 provides a detailed, ten-step approach to the process of OSCE design and implementation. Each step concludes with a list of best practices or guidelines. Table 1.2 compares some of the steps and best practice recommendations with those developed by ASPE [14].

Chapters 3 and 4 are devoted to the expansion of performance-based assessment as an educational methodology. Good OSCE data predictably identifies and indicates strategies for helping learners in need of remediation, as surveyed in Chap. 3. Looking beyond the training context, Chap. 4 explores how demands for more "in vivo" assessment can be met through the use and implementation of incognito or unannounced SPs (USPs) in clinical settings, while announced standardized patients (ASP) can enable workplace-based learning. Chapter 5, focuses on post-OSCE note assessment and clinical reasoning competencies such as information organization and synthesis and differential diagnosis and management plan. Chapters 6 and 7 highlight how to display data for learners and educational leadership and build, facilitate and expand scholarship at your own institution. OSCE are creative and team educational events. Chapter 8, showcases the full potential of performance-based assessment to address curricular and assessment needs. Chapter 9, examines the best practices for teaching and assessing telehealth skills in the post COVID-19 era. Chapter 10, focuses on the use of immersive simulations to prepare near graduates for their professional transitions. Chapter 11, reviews how to create teaching experiences/exams to strengthen clinical teaching environments. Chapter 12, describes how to create a sustainable and high-quality Standardized Patient Program at your own institution to support all your performance-based assessment activities. The Appendices at the back of this book contain blank versions of the forms and worksheets included in the main text, a sample OSCE and OSTE case and checklists, and resource suggestions.

Many institutions have developed sophisticated clinical skills centers. However, this is not essential for continuing to develop valuable learning opportunities and important methods for testing clinical skills. For many years, we produced OSCEs in empty classrooms or walk-in clinics on weekends, using carefully designed clinical scenarios and well-trained SPs. No matter how small or large your group of learners, this book can help you develop and expand your teaching and assessment armamentarium. While OSCEs are resource-intensive endeavors and require inspirational leadership to

Table 1.2 Cross-validation of the Chap. 2 steps and best practice recommendations with the Association for SP Educators Standards of Best Practice 1.0

Organizing OSCEs (and other SP exercises) in ten steps	ASPE Standards of Best Practice (SOBP) [14]
Step 1: Identify available resources **Step 2: Agree on format and timelines** **Step 3: Blueprinting**	**Domain 2. Case development**
• Build a team with a variety of skills	2.1.2 Identify and engage relevant subject matter experts to assist in the creation of materials
• Delineate core competencies • Align OSCE skills and content assessed with current or new curricula • Establish performance criteria for each level of training	2.1.1 Ensure that cases align with measurable learning objectives
Step 4: Case development	
• Choose scenarios that are both common and challenging presentations for your learners • Build specific goals and challenges into each scenario	2.2.1 Clear goals and objectives that can be assessed 2.2.2 Goals and objectives that specify the intended level of learners
• Ensure that cases represent the patient population in your clinical environment	2.1.3 Ensure that cases are based on authentic problems, and respect the individuals represented in a case to avoid bias or stereotyping marginalized populations
Step 5: OSCE rating forms	
• Develop rating items based on the blueprint, and ensure that a sufficient number of items are included to reliably assess competence within the targeted domains	
• Consider using both behavior-specific items and global rating items in OSCE rating forms to enhance objectivity in representing what happened during the encounter and provide learners with specific and more holistic feedback	2.2.9 Include: Evaluation instruments and performance measures (e.g., checklists and rating scales, participant, and facilitator evaluations)
Step 6: SP recruitment and training	**Domain 3. SP training**
• Explore the psychological and physiological impact a case has on the SP to avoid toxic side effects	3.2.2 Engage SPs in discussion and practice of role portrayal features 3.2.3 Provide SPs with strategies to deal with unanticipated learner questions and behaviors
• Practice all aspects of the encounter, and do not leave SP performance to chance	3.2.5 Ensure SP readiness for the simulation activity through repeated practice and targeted feedback
Step 7: Evaluator recruitment and training for rating and feedback tasks	
• Bring multiple evaluators together to jointly observe a learner's performance on tape or live, compare ratings, and discuss similarities and discrepancies. • Practice giving feedback • Make raters aware of potential biases and rating mistakes	3.4.6 In formative assessment, ensure consistent and accurate completion of an assessment instrument within individual SPs and among groups of SPs performing the same task
Step 8: Implement the OSCE: Managing the session	
Step 9: Manage, analyze, and report data	
Step 10: Develop a case library and institutionalize OSCEs	

OSCE Force

Behind every healthcare hero
there's an invisible army of educators
creatively orchestrating
the conditions for learning,
managing the many moving parts,
developing, adapting, implementing,
challenging their own conditioning,
staying open to new ideas,
anticipating the future, and
passionately living their calling
as they protect and progress
the professions that heal.

Fig. 1.3 The Good Listening Project sponsored by the AAMC. Listener Poet Jenny Hegland interviewed E Kachur on May 6, 2021, which resulted in the OSCE Force poem [30]

guide meticulous planning, preparation, implementation, and evaluation, the benefits to all involved make the investment well leveraged.

The process and implementation of a successful OSCE can result in a powerful synergy capable of invigorating educational programs. The event itself brings together faculty, learners, and staff to put their efforts toward a common goal. OSCEs produce meaningful experiences and useful data. Despite the enormous work involved, most people leave the event recognizing the value and feeling enriched. The poem in Fig. 1.3 illustrates the passion that can drive such endeavors [30].

References

1. Ba H, Zhang L, He X, Li S. Knowledge mapping and global trends in the field of the objective structured clinical examination: bibliometric and visual analysis (2004–2023). JMIR Med Educ. 2025;11:e71844. https://doi.org/10.2196/71844.
2. Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in clinical neurology. J Med Educ. 1964;39:802–5.
3. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J. 1975;1(5955):447–51. https://doi.org/10.1136/bmj.1.5955.447.
4. Hodges B. The objective structured clinical examination: a socio-history. LAP Lambert Academic Publishing; 2010.
5. Petrusa ER, Guckian JC, Perkowski LC. A multiple station objective clinical evaluation. In: Proceedings of the twenty-third annual conference on research in medical education; 1984. p. 23211–6.
6. Morrison LJ, Barrows HS. Educational impact of the Macy consortia: regional development of clinical practice examinations. Final report of the EMPAC project, 1998.
7. Yedidia MJ, Gillespie CC, Kachur E, et al. Effect of communications training on medical student performance. JAMA. 2003;290(9):1157–65. https://doi.org/10.1001/jama.290.9.1157.
8. Sutnick AI, Stillman PL, Norcini JJ, et al. ECFMG assessment of clinical competence of graduates of foreign medical schools. JAMA. 1993;270(9):1041–5. https://doi.org/10.1001/jama.1993.03510090025006.
9. Baker TK. The end of step 2 CS should be the beginning of a new approach to clinical skills assessment. Acad Med. 2021;96(9):1239–41. https://doi.org/10.1097/ACM.0000000000004187.
10. Katsufrakis PJ, Chaudhry HJ. Evolution of clinical skills assessment in the USMLE: looking to the future after step 2 CS discontinuation. Acad Med. 2021;96(9):1236–8. https://doi.org/10.1097/ACM.0000000000004214.
11. Touchie C, Pugh D. Cancel culture: exploring the unintended consequences of cancelling the Canadian national licensing clinical examination. Can Med Educ J. 2022;13(4):62–7. https://doi.org/10.36834/cmej.73889. Published 2022 Aug 26
12. Stillman PL, Brown DR, Redfield DL, Sabers DL. Construct validation of the Arizona clinical interview rating scale. Educ Psychol Meas. 1977;37(4):1031–8. https://doi.org/10.1177/001316447703700427.
13. Anderson MB, Kassebaum DG. Proceedings of the AAMC consensus conference on the use of standardized patients in the teaching and evaluation of clinical skills. Acad Med. 1993;68:437–83.
14. Lewis KL, Bohnert CA, Gammon WL, et al. The Association of Standardized Patient Educators (ASPE) Standards of Best Practice (SOBP). Adv Simul (Lond). 2017;2:10. https://doi.org/10.1186/s41077-017-0043-4. Published 2017 Jun 27.
15. Yusoff MSB. Multiple Mini Interview as an admission tool in higher education: insights from a systematic review. J Taibah Univ Med Sci. 2019;14(3):203–40. https://doi.org/10.1016/j.jtumed.2019.03.006.
16. ABMS. In: Holmboe ES, Iobst WF, editors. *Assessment Guidebook*. Accreditation Council for Graduate Medical Education; 2020.
17. Stewart MG. Core Competencies; 2001. http://www.acgme.org/acWebsite/RRC_280/280_coreComp.asp
18. Frank JR, Snell L, Sherbino J, editors. CanMEDS 2015 physician competency framework. Ottawa: Royal College of Physicians and Surgeons of Canada; 2015.
19. Zabar S, Hanley K, Kachur E, Stevens D, Schwartz MD, Pearlman E, Adams J, Felix K, Lipkin M Jr, Kalet A. "Oh! She doesn't speak English!" Assessing resident competence in managing linguistic and cultural barriers. J Gen Intern Med. 2006;21:510–3. https://doi.org/10.1111/j.1525-1497.2006.00439.x.
20. Aeder L, Altshuler L, Kachur EK, Barrett S, Hilfer A, Koepfer S, Schaeffer H, Shelov SP. The "Culture OSCE"—introducing a formative assessment into a postgraduate program. Educ Health. 2007;20:11.
21. Altshuler L, Kachur EK. A culture OSCE: teaching residents to bridge different worlds. Acad Med. 2001;76:514. https://doi.org/10.1097/00001888-200105000-00045.
22. Altshuler L, Kachur E, Krinshpun S, Sullivan D. Genetics objective structured clinical exams at the Maimonides infants & Children's hospital of Brooklyn. N Y Acad Med. 2008;83:1088–93. https://doi.org/10.1097/ACM.0b013e31818902e5.
23. Chander B, Kule R, Baiocco P, Chokhavatia S, Kotler D, Poles M, Zabar S, Gillespie C, Ark T, Weinshel E. Teaching the competencies: using objective structured clinical encounters for gastroenterology fellows. Clin Gastroenterol Hepatol. 2009;7:509–14. https://doi.org/10.1016/j.cgh.2008.10.028.
24. Parish SJ, Ramaswamy M, Stein MR, Kachur EK, Arnsten JH. Teaching about substance abuse with objective structured clinical exams. J Gen Intern Med. 2006;21:453–9. https://doi.org/10.1111/j.1525-1497.2006.00426.x.
25. Zabar S, Hanley K, Stevens D, Kalet A, Schwartz M, Pearlman E, Brenner J, Kachur EK, Lipkin M. Measuring the competence of residents as teachers. J Gen Intern Med. 2004;19(5 Pt 2):530–3. https://doi.org/10.1111/j.1525-1497.2004.30219.x.
26. Beltran CP, Wilhite JA, Hayes RW, LoSchiavo C, Crotty K, Adams J, Hauck K, Crowe R, Kudlowitz D, Katz K, Gillespie C, Zabar S, Greene RE. Practice makes perfect: objective structured clinical examinations across the UME-to-GME continuum improve care of transgender simulated patients. J Grad Med Educ. 2024 Apr;16(2):182–94. https://doi.org/10.4300/JGME-D-23-00573.1.
27. Beltran CP, Wilhite JA, Gonzalez CM, et al. Requested a different doctor: developing and evaluating an OSCE assessing core skills in supporting trainees facing patient discrimination. J Gen Intern Med. 2025; Jan;40(1):207–12. https://doi.org/10.1007/s11606-024-09021-0.
28. Harden RM, Lilley P, Patricio M. The definitive guide to the OSCE: the objective structured clinical examination as a performance assessment. Elsevier; 2016.
29. Cawthorpe D. AI-driven learning in medicine: revolutionizing Objective Skills Clinical Examination (OSCE) preparation through the lens of the International Classification of Diseases (ICD); 2023.
30. Hegland J. OSCE force. AAMC good listening project, vol 3; 2021. https://www.goodlistening.org/poetrycollections/association-of-american-medical-colleges-2021
31. Kachur E. OSCEs in the US. In Otaki J (ed). Theory and practice of the OSCE. Tokyo: Shinohara Shuppan Shinsha. 2007. p. 204–17 [Japanese].

Organizing OSCEs (and Other SP Exercises) in Ten Steps

2

Elizabeth Kachur, Sondra Zabar, Kathleen Hanley, Adina Kalet, and Colleen Gillespie

Step 1: Identify Available Resources

Assemble a Team

As with most other educational programs, OSCEs and other SP projects can be a major undertaking. While it is necessary to have strong leaders who believe in the benefits of such comprehensive assessment programs, many other individuals are needed for adequate planning, preparation, and implementation. Table 2.1 details the different roles that OSCEs typically require. Some people may be able to hold multiple roles (e.g., SP and rater) and some roles may be shared among several individuals (e.g., co-leadership). Some OSCE workers will be involved only in the implementation phase of the OSCE (e.g., raters). However, there will be a need for a "core team" (e.g., OSCE committee) that is responsible for planning and development in advance of the OSCE dates. Regularly scheduled meetings can help the committee become more established and help maintain the initial enthusiasm to participate in such a challenging, yet exciting undertaking. After the OSCE is completed, this group can work on data analysis and interpretation. Serious consideration should be given to sharing program experiences and scholarship at regional and national conferences and contributing to the OSCE literature. This will be of value for the OSCE committee members.

For those involved in the actual OSCE implementation, the most basic job requirements are availability, interest in the project and stamina. Two additional characteristics of great importance are precision and flexibility. Since OSCEs strive for standardization, it is necessary for all involved to be committed to keeping factors such as timing or case portrayal as consistent as possible. On the other hand, when dealing with large-scale events that involve so many people simultaneously, irregularities are likely to occur (e.g., a learner enters the wrong station, a rater arrives late). Thus, being flexible and willing to adapt is equally important.

It will not always be possible to find all the necessary players within your immediate work area. Thus, one should consider looking outside one's division and forging alliances across departments and levels of training (i.e., undergraduate, graduate, and continuing health professions education). Much of what is required for a successful OSCE is independent of specialty or profession.

Identify Location

By now, most institutions have created simulation centers, which are built for the explicit purpose of housing simulation events. These facilities usually include areas which are specifically designed for OSCEs by providing a series of exam rooms not much different from typical clinic settings. Usually, they are outfitted with one-way mirrors and/or video recording system that are operated from a control room. Some sim centers also include computers for learner's instructions and post-encounter exercises. Increasingly more often, SPs complete rating forms online using computers, iPads, or mobile phones, which adds further technologies.

COVID-19 forced many OSCEs to go virtual or else be eliminated altogether due to infection control concerns. As challenging as this transition may have been initially, it did turn out to be a silver lining of the pandemic. Over time, we

E. Kachur (✉)
Medical Education Development, Global Consulting, New York, NY, USA
e-mail: mededdev@earthlink.net

S. Zabar · K. Hanley
Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

A. Kalet
Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA

C. Gillespie
Institute for Innovations in Medical Education, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_2

Table 2.1 OSCE staffing needs (roles needed to run a smooth assessment program)

Roles	Key Characteristics	# Needed
Leader	• Strong motivation to develop and implement project • Well connected to procure resources, including access to institutional or local clinical skills testing facilities • Involved in curriculum decision-making • Able to communicate well, can negotiate and create a team spirit	0ne or more
Planner	• Understands logistics of implementing OSCEs • Is familiar with local conditions • Can entertain multiple options for solving problems	One or more
Administrator	• Can implement OSCE-related tasks (e.g., scheduling, SP recruitment, photocopying of station materials, online survey development, data entry) • Able to communicate well and create a team spirit • Good at troubleshooting and problem solving	One or more (depending on scope)
Station developer	• Has relevant clinical or content experience • Is familiar with performance standards • Accepts editing	One or more (depending on scope)
Trainer	• Understands SP and rater roles as well as case requirements • Has teaching skills (e.g., provides constructive feedback) and can manage the psychosocial impact of case portrayals • Able to communicate well and create a team spirit	One or more (depending on scope)
SPs	• Committed to standardization of their case portrayal (i.e., not expressing their personal creativity) • Comfortable enacting their particular case scenarios (i.e., not getting emotionally over-involved) • Interested in taking on "educational" responsibilities	At least one per station, consider cross-trained alternates
Rater	• Clear about OSCE goals and performance standards • Committed to fair performance assessments (e.g., understands personal rater style and biases) • Effective feedback provider (if learners receive post-encounter feedback)	At least one per station, consider cross-covering alternates
Timer	• Committed to maintaining the OSCE schedule • Able to focus despite periods of inactivity (e.g., when learners are in their stations) and distracters (e.g., SPs/faculty on breaks)	At least one
Monitor	• Able to direct rotation flow • Can troubleshoot and problem solve (e.g., faculty missing in station, lack of rating forms, video equipment problems)	At least one (may not be needed if institution has a dedicated clinical skills center)
Audio-visual technician	• Expert in using audio and video equipment (e.g., setup, recording, archiving) • Able to manage online program functionalities (e.g., breakout rooms, announcements) • Can troubleshoot technical problems (e.g., Wi-Fi or equipment issues)	At least one (may not be needed if institution has a dedicated clinical skills center)
Data manager	• Can enter performance data (e.g., coding, missing data management) • Understands OSCE process • Committed to accuracy	At least one
Data analyst	• Understands OSCE process • Has psychometric skills • Understands end-users of results (e.g., learners, program)	At least one
Program evaluator	• Understands OSCE process • Is familiar with evaluation models (e.g., pre/post testing) • Can develop and analyze program evaluations (e.g., surveys, focus groups)	At least one

learned to create OSCEs that are implemented via teleconferencing systems, and our OSCE site options expanded exponentially. In addition to an all-virtual event, there is also the possibility to create hybrid or blended OSCEs. In the latter, some stations occur online (e.g., a telemedicine encounter, see Chap. 9) or some participants in all stations are participating virtually (e.g., SPs). Of course, this adds technical personnel and equipment as well as a stable internet connection to the list of essential resources.

When planning an OSCE, most organizers will aim for an environment that closely resembles real clinical settings. Ambulatory care settings are the ideal alternative space for OSCE stations, and one can schedule them during weekends or other times when the clinic is closed. A close coordination with clinic administrators and personnel before, during, and after the event will be critical. But, while authenticity is important, availability is essential. To facilitate the demonstration of true competencies and/or enhance transfer from

the simulation to the real-life setting, matching the environment to the tasks in focus is not all that is needed.

Organizations such as the Society for Simulation in Healthcare (SSH) [1] and the Association for SP Educators (ASPE) [2] have a myriad of rich resources such as conferences, webinars, and listservs, to provide information about physical setups, online platforms, videotaping, and data management companies. Regardless of how up-to-date and luxurious the available setups may be, there is a lot more needed to create successful learning and assessment environments. With enthusiasm for the program, creativity, and commitment to provide a quality experience for all involved, even an empty seminar hall or classroom can result in a viable and effective educational program.

Identify Sources of Funding and Support

There are many venues to explore for funding SP activities and pilot programs. Begin by investigating your own institution's medical education resources at the level of the dean's office, department, and division. In addition, some SP programs have been funded by local medical societies, foundations (e.g., through grants for improving doctor-patient communication and safety), and philanthropic sources.

Best Practices: Assembling a Team

1. Build a diverse team which is enthusiastic about OSCEs.
2. Assure that the team members have the needed skill sets or are willing to acquire them.
3. Schedule regular meetings to build group identity.
4. Create a common repository (i.e., shared drive, secure website) for meeting minutes, materials, and protocols.
5. Look broadly for suitable sites (physical or virtual spaces) and consider hybrid OSCE formats.
6. Explore potential funding sources.

Step 2: Agree on Formats and Timeline

Once the decision is made to organize an OSCE, further details need to be worked out. A worksheet such as that shown in Table 2.2 (also included in blank form as Appendix A at the back of this book) can assist with this task. It is often necessary to balance educational opportunities with available resources and strategic considerations.

From the start, it has to be clear whether the OSCE is meant to be formative or summative. Formative OSCEs are teaching exercises where the focus is on experiential learning that is enhanced by feedback and debriefing. Summative OSCEs are meant to support admission or promotion decisions, such as

Table 2.2 Example worksheet for making initial OSCE plans

OSCE project name	*Annual General Internal Medicine Residency OSCE*
OSCE goals	*Assessment of general clinical competencies*
Number and type of trainees	*20 PGY1-3s*
Number and type of stations	*10 stations = 8 independent stations with SP encounters; 2 rest stations*
Number of SPs needed	*11 SPs = 9 primary SPs, 2 back-up SPs*
Number of faculty needed	*9 faculty = 1 faculty to lead prebrief/debrief, 8 faculty observers*
Potential timing	*All residents on 1 day (two 3.5-hour sessions)*
Potential space	*Simulation center/Outpatient department with no patient care sessions*
Approximate budget	*$4500 (SPs, food)*
Potential funding sources	*$3000 Residency budget* *$1500 Institutional grant*
Instructional strategies	*Group pre-brief/orientation; individual immediate post-station feedback from faculty observers and SPs, group post-OSCE debrief; individual resident and aggregate education leadership reports*

passing a course or being ready for the next phase of training or graduation. In reality, the division between these two types is not always clear cut. Many programs use formative OSCE data in performance reviews (e.g., competency committees) to determine remediation needs. Thus, it is more a continuum rather than two discrete entities.

The OSCE format also needs to be determined early on: will it be implemented in-person, online, or hybrid? There are benefits and drawbacks to all modalities as outlined in Table 2.3. Many programs have already acquired much experience with in-person OSCEs, but due to COVID-19, there has been a sharp rise in online OSCEs. With increasing familiarity in both formats and the definite change in medical practice patterns, hybrid OSCEs are likely to dominate in the future.

Table 2.4 provides a list of core OSCE budget items, filled in for the same example 8-station General Internal Medicine Residency OSCE introduced in Chap. 1 Fig. 1.2. SP training time accounts for case portrayal and checklist rater training and includes 2 extra SPs for back-up; cost per learner is calculated for 20 residents. A blank version of this budget form is also included as Appendix B to assist readers in making cost and resource projections. With most projects, funding will be of concern. However, there are various ways to manage with fewer resources. For example, by adding 2 rest stations to an 8-case OSCE, it is possible to include all 20 residents in just two 10-station OSCE runs, thus saving an entire additional OSCE day. Other saving strategies include using staff as SPs (though not ideal) or delegating assessment and feedback responsibilities to trained SPs to reduce faculty participation needs [3–5].

Generally, one is wise to start small, and then expand to more complex and ambitious training or assessment programs. By beginning with a pilot project, one can develop local expertise and generate enthusiasm among learners and teachers. Formative assessments that focus on learning will require fewer resources and demand less stringency regarding case portrayal and rating accuracy than high-stakes exams. They are likely to be less stressful for all involved, and thus have a better chance to convert skeptics.

Table 2.5 shows a worksheet used in planning for our example General Internal Medicine Residency OSCE to assign tasks and prepare a project timeline (a blank copy of this worksheet is also included as Appendix C). Typically, one needs to start work 3–4 months before the event. However, with the help of individuals who already have

Table 2.3 Benefits and drawbacks of OSCE formats: in-person, hybrid, and online

	In-Person	Hybrid	Online
OSCE Design	All stations are onsite, all learners, SPs, and faculty participate in-person	• Some stations or some participants attend in-person, others are online (e.g., telemedicine OSCE station, telephone consultation)	• All stations, learners, SPs, and faculty participate online (e.g., via a teleconferencing platform)
Opportunities	• Well-established format, individuals with extensive experience are easier to find • Facilitates inclusion of physical exams, procedures, task trainers, and high-fidelity mannequins • Physical setting resembles frequent work environments (e.g., clinic exam room) which enhances validity of the assessment and transfer of newly acquired skills • Rating forms and program evaluations can be collected onsite or checked immediately online before everyone leaves allowing for more data collection control • Physical proximity facilitates control of all participants and allows for easy interventions if problems occur	• Reduces the need for dedicated physical space • Flexibility with SP recruitment (e.g., better chance to find SPs with the "right" fit) • Facilitates last-minute SP substitutions • Easier to schedule training (in-person and online OSCE can train virtually) • Learners and faculty at remote locations do not need to travel to the OSCE site which increases efficiency and decreases costs • Approximates future hybrid practice patterns • Allows for telemedicine-type encounters • Optimal flexibility for creating a mix of stations to meet curriculum needs	• No need for physical space allocations (e.g., independent of Simulation Center availability) which provides more scheduling flexibility • Possible prior experience with format during COVID-19 pandemic • Can teach/assess competencies during telemedicine encounters • Easier to recruit SPs • Training can be done online for all cases and may be easier to schedule • Learners, SPs, and faculty save time since they do not have to travel to the OSCE site • Potential cost savings (e.g., no need for refreshments, no travel or parking costs) • Online platforms (e.g., Zoom) provide video recording options
Challenges	• Need for appropriate physical space • Need for SPs who are willing and able to work onsite • Does not account for changes in practice patterns (e.g., telemedicine) • May require additional expenditures (e.g., refreshments, parking tickets)	• Requires physical and online setup (e.g., stable internet connection and backchannel) • Unless additional staff is available, more need for multi-tasking (e.g., to coordinate timing, announcements, and support for both formats) • Requires strategies to collect evaluation data in both modalities	• Requires stable internet connection • Requires a reliable communication backchannel for coordination • Timing and announcements can be more difficult • Rating forms may require separate monitors/devices or paper forms that are transcribed afterward

Table 2.4 Example OSCE budget

Budget items to consider	Amount needed	In kind	Cost/learner
Space (SIM center or true clinical space) 1 room per station, SP/faculty and learner meeting areas		SIM center facility fees are waived for GME	
SPs—training and OSCE Performance Check for local rates, costs vary depending on location and simulation task. Factor early arrival and debriefing/deroling time into payment	$28/hour x 11 SPs x 12 hours (4 hours training + two 4-hour OSCE sessions) **= $3,696**		**~$185**
Faculty observers (feedback and prebrief/debrief)		Donated	
Medical supplies Does not need to be sterile, but should be authentic		Donated	
Office supplies Printing paper evaluation forms, pens		Donated	
Refreshments for learners, SPs, and faculty	**$120**		
Recording equipment		Provided by SIM center	
Data entry and report assembly (individual learner data and aggregate program data) May be performed via LMS[a] or staff		Part of SIM Center's LMS[a]	
Data analysis Faculty/staff with statistical analysis capabilities are vital for scholarship and longitudinal tracking	$30/hour x 20 hours = **$600**		**$30**
Total	**=$4,416**		**=$215**

[a]Learning Management System

	Less Challenging	More Challenging
Data Gathering	• History of present illness • Past medical history • Psychosocial history • Occupational/environmental history	• Sexual history • Substance abuse history • Suicidal ideation • Domestic violence • Cultural/religious practices • Mental status exam
Patient Education & Counseling	• Simple issues • Aligned health beliefs • Motivated patient	• Multifaceted problems • Nonaligned health beliefs • Resistant patient
Negotiations & Shared Decision Making	• Agreement between parties • Understanding	• Disagreement between parties • Lack of understanding

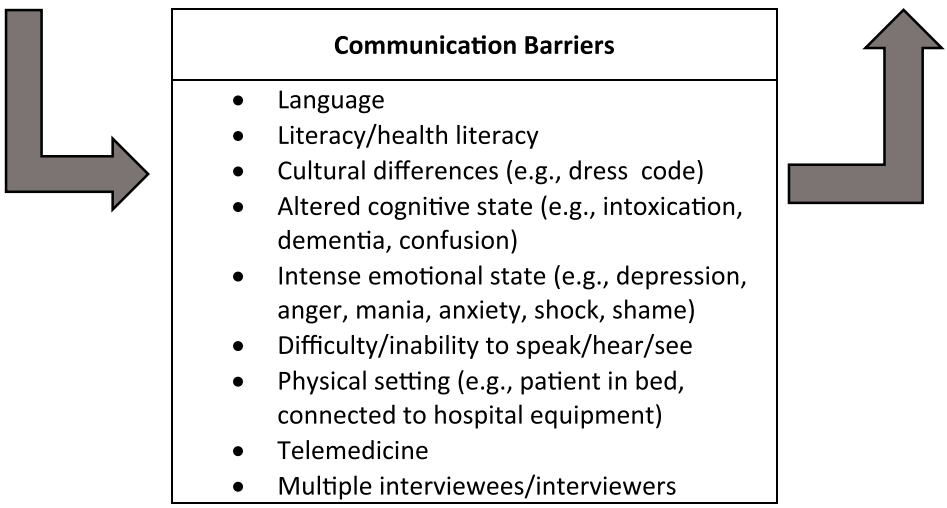


Fig. 2.1 Adjusting case difficulty

Table 2.5 Example worksheet for assigning OSCE responsibilities and creating timelines

OSCE Project Name: *Annual General Internal Medicine OSCE*		Date: *March 19*	
Tasks		Individuals involved	Deadlines
Initial planning			
3–4 months before the OSCE	Decide on format (e.g., time frame, modality, number of stations)		*Nov 24th*
	Create a blueprint (identify competencies to be assessed)		*Nov 24th*
	Identify appropriate OSCE location (stations and assembly rooms) or online platform		*Dec 17th*
	Recruit staff (for administrative tasks, monitoring, time keeping)		*Dec 17th*
	Communicate with learners (provide dates/times, explain format and procedure)		*Dec 17th*
	Clarify and negotiate budget (e.g., SP costs, refreshments)		*Dec 17th*
	Consider videotaping and arrange for set-up		*Dec 17th*
	Decide on what stations to maintain from previous OSCEs/develop new/import from station banks		*Jan 1st*
	Decide on SP/rate recruitment and schedule		*Jan 7th*
Station/material preparations			
1 week – 3 months before the OSCE	Determine SP payment process		*Dec 17th*
	Make room arrangements and/or determine online platforms		*Dec 17th*
	Recruit faculty for pre-brief, debrief and observation		*Jan 14th*
	Prepare station materials (i.e., develop new ones, adapt old ones if appropriate)		*Jan 31st*
	Recruit SPs		*Feb 18th*
	Prepare faculty (e.g., circulate station/format information, feedback training)		*Feb 18th*
	Prepare props (e.g., fake pill bottles, charts)		*March 4th*
	Train SPs		*March 11th*
General on-site and/or online preparations			
1–2 weeks before the OSCE	Order supplies (e.g., paper, folders)		*March 4th*
	Assign SPs, faculty and learners (create assignment sheets)		*March 4th*
	Develop rotation schedules (i.e., matrix of learners, rotations and stations; if online, it will be useful to include pictures of learners to double check that the right person is in the right station)		*March 4th*
	Prepare invoices and necessary paperwork for SP payment		*March 4th*
	Print station materials		*March 11th*
	Prepare name tags/labels for learners (assign numbers to learners)		*March 11th*
	Prepare signs (e.g., station numbers, arrows to signal flow)		*March 11th*
	Orient hall monitors and timekeepers		*March 18th*
	Practice walk-through		*March 18th*
OSCE Administration			
Day of the OSCE	Prepare stations and assembly rooms (e.g., signs, station materials, refreshments) or set up breakout rooms on online platform (e.g., stations, SP, faculty and learner assembly rooms, orientation and debriefing rooms)		*March 19th*
	Assign substitutes (if necessary)		*March 19th*
	Orient faculty, SPs, and other personnel (in-person and/or online)		*March 19th*
	Position faculty, SP, hall monitors, timekeepers (in-person and/or online)		*March 19th*
	Orient learners (in-person and/or online)		*March 19th*
	Guide learners to individual starting stations		*March 19th*
	Time stations and make announcements (start, feedback, station changes, end of OSCE)		*March 19th*
	Manage emergencies (e.g., equipment/Wi-Fi breakdown)		*March 19th*
	Assure smooth changeovers of SPs, faculty, learners (in-person and/or online)		*March 19th*
	Assemble learners for the OSCE group debriefing (include faculty if possible)		*March 19th*
	Assemble SPs for debriefing and de-roling		*March 19th*
	Collect and count all forms (check completions if submitted online)		*March 19th*
	Clean up stations and assembly rooms		*March 19th*

(continued)

Table 2.5 (continued)

OSCE Project Name: *Annual General Internal Medicine OSCE*		Date: *March 19*	
Tasks		Individuals involved	Deadlines
Post-OSCE Tasks			
Days to weeks after OSCE	Debrief with organizers (if possible, include timekeepers, SP trainers and others involved in the administration)		*March 19th or later*
	Sort out forms (in-person and/or online)		*March 21st*
	Ensure timely SP payment		*March 25th*
	Enter data and evaluation results if not completed through an electronic Learning Management System		*March 25th*
	Analyze data		*April 8th*
	Report evaluation data (e.g., report cards)		*April 19th*
	Report on experience internally and externally (e.g., presentations, articles)		*Sept 19th*

expertise in this area and the availability of OSCE case banks, shorter planning times may be possible.

Best Practices: OSCE Planning

1. Identify date and time of the OSCE (consider potential conflicts with holidays, conferences, and/or other programmatic events).
2. Determine the OSCE format (i.e., In-person, Hybrid, Online).
3. Secure participants' availability.
4. Identify potential physical location and/or online platform to be used.
5. Start early to identify potential SPs and faculty needs.
6. Make a timeline working backward from the OSCE date.
7. Make a budget at the outset of the project.

Step 3: Identify Potential Stations and Establish a Blueprint

A key element for designing an OSCE is the development of a blueprint. This is a matrix that connects a list of all potential or planned stations with the competencies that are to be assessed. Table 2.6 provides an example; a blank matrix is included as Appendix D. A blueprint ensures that individual competencies are examined multiple times and that each station assesses multiple competencies. In doing so, one can verify the overall comprehensiveness of the exam or exercise.

By drafting a blueprint when brainstorming what type of cases to include, one can easily ascertain which content areas are overrepresented and what type of scenarios are still needed. Thus, one can arrive at a comprehensive formative or summative assessment tool. Some stations may have to be created, others could be adapted from other programs (e.g., from MedEdPORTAL [6]), or they could be recycled from one's own collection. A case bank like the one described in Step 10 (see Table 2.22) is a useful repository which organizes accumulating cases by content, SP characteristics, nature of encounter, and skills covered. A case library can also track station usage to facilitate competency tracking over time/cohorts or to avoid overexposure of specific stations. By including literature reviews or curriculum surveys, one can illuminate the need and utility for using specific cases. Station-specific readings or other resources can help deepen post-OSCE debriefings and remediation efforts.

An organized approach to blueprinting strengthens an OSCE's validity. Whether they are formative or summative, OSCEs should provide a good cross section of medical encounters typically experienced by learners. Post-OSCE feedback from trainees (see Appendix K for a participant post-OSCE survey) should confirm that the stations assess issues they encounter in their current studies or that they are likely to face in their future practice.

Once a first draft of a blueprint is completed, organizers should ask themselves the questions listed in Table 2.7. The final sequence of stations is guided by several considerations, including variability of case gender and emotional tone (e.g., two "angry patient" stations should not be next to each other) as well as site or station limitations (e.g., only certain rooms have an external phone connection).

Best Practices: Blueprinting

1. Delineate core competencies/skills to be covered and assure that they are aligned with current or new curricula.
2. Each competency should be assessed by multiple stations and each station should cover more than a single competency.
3. Consider specific competencies (e.g., communication skills) to be tested in all stations.
4. Ensure that the station characteristics (e.g., patient age, gender, race, prevalence of disease) reflect current or future actual clinical practice.

Table 2.6 Example OSCE blueprint

Station			Competencies					Implementation notes	SP characteristics
Case number and name	**Case objectives**	**Content areas**	**Core Communication Skills**	**History Gathering**	**Education & Counseling**	**Treatment plan & Management**	**Case-Specific Skills**		
1. Urethritis follow-up	Take sexual history from trans patient	Prevention (HIV, sexual History, STI prevention)	X	X		X	Health equity	Opportunity to connect with LGBTQIA+ - acting group	Trans woman SP, 20s, Latina
2. High-risk smoker	Move patient from contemplation to action stage, develop plan	Addiction medicine (smoking, behavior change cycle)	X	X	X	X	Counseling		Male SP, 50s, Black
3. Supporting a resident experiencing bias	Support intern who is rejected by a patient	Team communication/ support	X	X		X	Allyship	Opportunity to explore other biases in debrief	Female standardized learner (SL), 30s, wears Hijab
4. New diagnosis: atrial fibrillation	Educate patient on new onset "palpitation" and treatment	Pt education regarding: blood thinner; risk/benefit and teach back	X	X	X	X	Shared decision-making	Make sure learners are up to date with current medication recommendations	Female SP, 50s, white
5. Difficulty sleeping	Screen appropriately and identify behavioral health issues	Common symptoms/ undifferentiated problems (PTSD, Diagnosis of sleep problems)	X	X		X	Use of screening tools	Make sure SP aligns with screening tools for PTSD and depression	Male SP, 30s, white
6. Diabetes precepting	Precept standardized learner on new onset diabetes case	Acute problems (diabetes), microskills teaching	X		X	X	Teaching		Male Standardized Resident (SR), 30s (consider using medical student), Black
7. Post-discharge televisit	Call post-discharge patient with chronic heart failure	Facilitate physical exam though screen, provide transition of care	X	X			Virtual health	Patient starts encounter with camera turned off, reports swollen legs upon guided virtual exam, has a list of medications to show	Male SP, 50s, Asian
8. Inter-professional case	Manage acute chest pain with a nurse	Interprofessional acute chest pain management	X	X	X	X	Closed-loop communication	EKG alternates between right-sided MI vs. aortic aneurysm; props needed (EKG, medications, algorithm for standardized nurse)	Male Standardized Nurse (SN), 30s, white Female SP, 60s, Black

Table 2.7 Important questions for blueprint development

☐ Are cases representative of typical clinical practice?
☐ Are cases representative of what has been taught in the course/rotation?
☐ Do the cases adequately cover all the competencies to be tested?
☐ Are diagnostic and management challenges varied in a systematic fashion?
☐ Is there a balance in terms of gender, either equally divided or resembling real-life practice?
☐ Is there an appropriate mix of patient ages?
☐ Is there an appropriate mix of racial and cultural backgrounds?

Step 4: Develop Case Scenarios and Instructions

The blueprint leads to profiles for each of the stations which then can serve as starting points for case development (the *case* is the clinical, professionalism or teaching problem, which involves a specific set of tasks to be assessed or taught). Basing OSCE stations on real scenarios and challenges will add validity. However, after masking the identity of the persons involved, it may also be necessary to make adjustments for the training level, OSCE focus, or the time limitations imposed by the exercise. Figure 2.1 illustrates how one can modulate the difficulty level of communication tasks. By making stations more or less challenging, one can also increase or decrease the overall difficulty of the OSCE.

Our case development worksheet is included in Appendix E. Our outline follows a variety of clinical encounter frameworks and is compatible with the 2017 ASPE Case Development Template [7]. The focus on a patient encounter can easily be adjusted to scenarios that involve family members, learners, or other healthcare providers. See also Appendix F for a rating form development worksheet.

Educators should not feel obligated to start from scratch in developing their OSCE cases. Additional station and case development resources are included in Appendix P. We also recommend reaching out to other health professions schools; many programs will likely be willing to share their OSCE cases. Usually, some adjustments will be necessary to make the station relevant to one's learners and local circumstances, but it definitely helps to get a quick start.

Station development is a process which does take various iterations, following the principles of "emergent design." Drafting, reviewing, editing, and piloting prior to implementation require time which needs to be incorporated in the overall plan. The ASPE Standards of Best Practice [8] also place much emphasis on the importance of co-creation. By incorporating the thoughts and experiences of many stakeholders (e.g., content experts, SP educators, patients), one has the best chance to create an OSCE station that is feasible, of high fidelity and meaningful considering its purpose. Collaboration can involve interviews and focus groups with patients or other target groups (e.g., transgender populations), roleplay with OSCE Committee members to identify key competencies to assess or address timing issues, and adjustments made during SP training to make sure that the prescribed wordings and emotional tones appear natural and can be standardized. For example, a surprise crying spell during an encounter may not just be exhausting for the SP, but it will also leave physical traces on his/her face which will be visible to the next learner from the start. Consequently, the fidelity of the case will suffer.

Case materials for the SP and faculty need to be sufficiently detailed to assure consistency. Yet, they must not be so voluminous that there are too many details to remember and to reproduce consistently. Regardless of preparation and training time, cognitive overload is never helpful. Table 2.8 provides considerations specific to each component of the materials accompanying each station. In addition to written materials, one can also consider other pre-station strategies such as including a brief video or mannequin task that "sets the scene" in terms of content and emotional tone (e.g., previous observation of unprofessional behavior, an unsuccessful procedure performed on a mannequin). Case materials for an example case, "Medical Error Disclosure," including station overview, directions for the OSCE participants, and detailed SP case portrayal instructions, can be found in Appendix G. Corresponding rating forms for both the SP and faculty observer are provided in Appendices H and I, respectively

Each OSCE form should be clearly marked with station number and title. Institutional origin, date stamps, and confidentiality requests can be added in headers or footers. The title needs to be phrased in order not to give away the sometimes hidden, station-specific challenge (e.g., do not use "Secret Drinker"). Some OSCE organizers prefer to use the name of the person the learner will encounter in the station (e.g., the patient's name).

In designing a scenario, one should also consider how to use the time immediately following the SP encounter. Post-encounter options for the learner include writing up a patient note, interpreting additional diagnostic information or presenting the case, thus making this a "paired station." Reflection and instant feedback after an encounter are especially important for formative OSCEs and typically appreciated by learners [9]. Which option one selects will depend on one's goal for the OSCE and the station, as well as pragmatic considerations such as timing.

If learners receive feedback after each encounter there are typically strict time limits. Thus, it is very important to provide clear guidelines for the observer, whether it is a faculty member or the SP. Table 2.9 provides a sample set of instruc-

Table 2.8 Overview of station-specific materials (their purpose, content, and special considerations)

Forms	Purpose	Content Elements	Considerations/Tips
Station overview	To assist program organizers	• Date of development/1st use • Station Developers/Institutions • Learners—intended and potential (type, training level) • Station goals/objectives (what is the purpose of this station) • Competencies to be assessed • Logistics (personnel, station materials, room arrangements)	• Be specific • Identify room requirements (e.g., telephone access) • Indicate updates/versions • Add station sources (e.g., MedEdPortal reference) if applicable
Learner instructions	To communicate the scenario and tasks to learners before they enter the station	• Patient/encounter information (e.g., name, age, occupation) • Reason for visit/encounter • Learner role • Starting point for encounter (beginning, middle, end) • Situation (medical/psychosocial information available, prior developments/encounters) • Learner task(s)	• Be brief and bullet information if possible (consider reading time) • Assure equal length with other stations • Timeline with arrows can help orient learner quickly • Adjust information to who learners will encounter (e.g., family member, healthcare provider) • Use language learners are familiar with (e.g., well-known abbreviations) • Consider brief videos to prime learners, quickly communicate complex content and emotions • Indicate if certain maneuvers are NOT to be performed (e.g., rectal exams)
Fact sheets (if applicable in selected stations or OSCEs)	To provide learners with information needed for managing the case if specific information is not familiar or if one tries to focus encounter on communication skills and wants to equalize the required medical knowledge	• Case-specific guidelines for diagnosis and treatment • Case-specific screening tools (if they would be present in a clinical setting) • Administrative or legal factors relevant to the case • Community resources	• Be brief (reading time is limited) • Organize material to be reviewed quickly • Use graphs where possible • Assure accuracy • Avoid controversy • Assure parity with other stations
SP instructions	To prepare SPs for their case	• Scenario (what happened from the SP's perspective, why is he/she here today, prior medical encounters) • Current life situation and past history (medical and psychosocial) • Personality and emotional tone (how to relate to the learner) • Cues to give to learners (verbal, non-verbal) • Timing (beginning, middle, end/after 2-minute warning)	• Provide opportunity for SPs to personalize scenario within limits (e.g., name of spouse) • Supply an "opening line" and "required statements" to provide learners • Balance level of detail (i.e., not too little, and not too much) • Illustrate the emotional tone to be portrayed with sample statements • Clearly identify specific challenges learners should manage

(continued)

Table 2.8 (continued)

Learner assessment checklist or rating form	To assess learner's competencies/behaviors and capture the performance assessments	• Paper or online format • Administrative information (e.g., learner IDs, date, station) • Dimensions on which to assess the learner (e.g., communication skills, case management) • General items (to be used across all stations) and station-specific items (tailored to the scenario) • Checklist or global rating items • Room for comments (e.g., areas of strengths, areas in need for improvement)	• Make items evidence-based • Adjust items to fit the type of encounter (e.g., family member, healthcare provider) • Keep the number of items manageable for the allotted rating time and for the ability of average raters to focus on during the encounter • Watch out for double negatives • Pre-test for readability and ability to observe and rate • Include at least one summary rating for cross-validation • Consider using station objectives as global rating form items • Consider paper versus computerized formats. If using online surveys, have backup paper checklists in case of technological difficulties
Faculty instructions	To standardize faculty assessment and teaching	• Procedural steps for observing encounters (e.g., positioning to observe non-verbal behaviors, start/stop video recordings) • Procedural steps for providing feedback (e.g., start with learner's self-assessment, invite SP to comment) • Teaching points (i.e., what messages to deliver to each learner if instant feedback is provided)	• Keep it brief • Use bullets when possible • Assure that procedures are consistent at all stations • Match teaching points with station objectives
Post-encounter materials (optional)	To give learners the opportunity to reflect on/synthesize the encounter, extend their clinical reasoning or receive feedback about the case or their performance	• Patient note (with space for summarizing history, diagnosis and treatment plan) • Supplementary diagnostic test results (e.g., EKG, X-ray) • Instructions for reporting on the previous encounter • Self-reflection surveys	• Be selective and pragmatic: e.g., weigh faculty availability for giving feedback vs. gathering further learner data • Consider computer- vs. paper-based options

tions that could help structure a brief feedback session. It will also be important to add 2-4 station-specific teaching points to make sure that the teaching objectives for each station can be covered routinely. See the feedback training protocol (Step 7, Table 2.16) below for some general considerations.

To assure the quality of each case, organizers should ask themselves the questions listed in Table 2.10.

Table 2.9 Sample framework for giving brief instant feedback during the OSCE

1.	**Start by asking the learner, "How did it go?"**
2.	**Reflect back key points**
3.	**Ask SP(s) for feedback** (if SP is not the only person giving feedback)
4.	**Ask the learner what was done well** Be prepared to discuss 1 item from the rating form Must be a specific behavior
5.	**Ask the learner what could be done differently** Be prepared to discuss 1 item from the rating form Must be specific behavior
6.	**Feed-forward: "**The next time you see a patient like this, what will you do?"

Best Practices: Case Development

1. For summative OSCEs choose scenarios that are both common and challenging for your learners. If the case is too easy, it will be difficult to discriminate learners based on their performance.
2. For formative OSCEs, select challenges learners are likely to face in the future (e.g., patient population, interprofessional contexts) or those which are critical but not covered locally due to gaps in the curriculum (e.g., lack of specific patient populations).
3. Build specific goals and challenges into each scenario. Also include social determinants of health and public health concerns.

Table 2.10 Review questions important for case development or adaptation

☐ **Are the station goals clear?** Do they provide precise information about what the station is supposed to teach or assess in terms of what learners need to know, be able to do and value?
☐ **Is the case appropriate for the learner?** Consider profession, training level, course/rotation content
☐ **Can the required tasks be successfully completed given the rotation intervals?** Given the allotted time (e.g., 10 minutes) would a skilled learner at the desired level be successful in demonstrating competence?
☐ **Are the learner instructions clear?** Can someone quickly ascertain what the situation is and what needs to be done? Are the instructions uniform across cases in terms of format and length?
☐ **Are the SP instructions clear?** Do they provide adequate background information for an SP to take on the role? Do they clearly indicate the key elements of the case, what is essential in terms of content, emotional tone, and timing?
☐ **Are the faculty instructions clear?** Do they provide adequate guidelines on how the faculty should proceed? Do they include appropriate, station-specific teaching points if post-encounter feedback is involved?
☐ **Is it possible to simulate the physical and/or psychological signs and symptoms for the length of time allocated to each rotation?** Can someone stay that depressed for 10 minutes? Will the case require multiple SPs because it is too stressful or too difficult to maintain a particular physical finding?
☐ **Do the case materials include elements that could result in stereotyping, creating, or perpetuating bias?** While some cases may need to highlight certain features (e.g., body characteristics, behaviors, dress code), is the wording such to support not distract from the station objectives?

4. Choose a meaningful post-encounter activity (i.e., feedback, supplementary exercise) if appropriate.
5. Routinely check for stereotyping and the bias-promoting case materials.
6. Make sure it is possible to complete tasks in the time allotted. A role-play with a skilled learner who is blind to the case is key for successful case development.
7. For summative assessments organize a trial run with learners who will not participate in the OSCE (e.g., from a different training program) to validate and fine tune cases.

Step 5: Create Rating Forms

The quality of a rating form is based on the degree to which it accurately captures the intended competencies or practices (validity). In order to be accurate, the rating form should first be consistent, that is, lead to the same results regardless of who the rater is (i.e., inter-rater reliability) or when the evaluation is performed by the same rater at different times (i.e., intra-rater reliability). Only once consistent can the validity of the assessment be determined. The keys to developing reliable and valid rating form items are a) identifying the specific knowledge, skills, and attitudes to be addressed, b) writing items that are unambiguous and easy to understand, and c) providing anchors or instructions that guide raters in their assessment. A blueprint (described in Step 3) specifies what skills and content the OSCE is designed to cover and how each station contributes towards this goal. Following the blueprint, rating forms need to include competencies that are assessed across all stations (e.g., communication skills), as well as content and skills specific to each station or subset of stations (e.g., patient education about medication).

The two main item formats are behaviorally anchored descriptions (e.g., *Did the trainee perform a specific behavior?* No/Yes or Not Done/Partly Done/Well Done) and global ratings (e.g., *How would you rate this trainee's professionalism?* Not at all professional/Somewhat professional/Professional/Very professional, or *Would you recommend this trainee?* Not Recommend/Recommend with Reservations/Recommend/Recommend Highly). Both types are important [10] and most rating forms include both. Each one can provide unique information about trainee performance, and often they are strongly correlated. If performance feedback is an important OSCE goal, behaviorally specific checklist items can provide learners with actionable data, while global ratings are much less helpful to guide learners toward improvement. Creating an adequate space for comments on the global ratings is particularly important. They can provide additional information about the learner's performance or recommendations on what could be done better [11]. They also offer an opportunity to indicate potential rating challenges (e.g., could not observe). Sample rating forms can be found in Appendices H and I.

Appendix F provides a rating form development template which includes a collection of items that cover a broad range of competencies, from communication and relationship building behaviors to physical exam maneuvers. Among the benefits of reusing items in different stations are ensuring a more robust, generalizable measure of the competence or behavior, identifying station-specific context influences, reducing training time because of item familiarity, and being able to track performance over time (development of competence).

Rating forms can be provided on paper and some programs use "no carbon required" (NCR) stocks to supply instant copies for immediate distribution to learners. With

the advent of survey software, widely available data collection platforms, and simulation software systems, most programs use technology-based rating forms. This requires some work upfront as forms need to be programmed and raters need to be trained to access these online tools appropriately. The major time savings and benefits of electronic rating forms for the program will come at the end, when analyzing and reporting the data. Digital data collection can improve completeness of data by requiring that items be filled out and including skip patterns and internal logic. By eliminating costly data entry, one can also forego potential transcription errors.

Checklist Items

Table 2.11 provides a stepwise process for developing behavior- or content-specific checklist items. Determining the number of items for addressing a particular domain (e.g., information gathering, relationship development, patient education and counseling, shared decision making, etc.) requires a good balance between coverage (i.e., validity) and burden on the rater. Asking evaluators to rate too many and/or very complex aspects of performance can lead to decreased accuracy and reliability. The classic study of SP ratings by Vu and colleagues identified 15 as the most optimal number of items that can yield an accuracy of 80.12% while covering a broad enough range of skills to satisfy validity needs [12]. Extensive, targeted training of raters, explicit behavioral anchors and adequate time for the rating tasks are additional ways to achieve a good balance.

Binary checklists are popular in OSCEs because of their simplicity—they can cover communication and procedural skills as well as content attainment specific to each station. They also focus on observation and recording rather than requiring judgments. This has been used as argument for letting SPs without clinical training evaluate clinical skills (e.g., physical exam maneuvers), essentially making them "reporters" rather than "evaluators." However, such simplification may miss important dimensions of performance and could, in some circumstances, compromise the validity of the assessment tool. Furthermore, such formats have been proven to not capture increasing levels of expertise since more advanced clinicians can take shortcuts without sacrificing diagnostic accuracy or effectiveness [13]. In addition, many raters object to simple yes/no checklists because so much of the behavior they witness falls into an area between those dichotomies. Consequently, scales that provide multiple rating options (e.g., Likert-type or even-numbered, forced choice formats) are often preferred. While more response options offer raters more opportunities to report on fine nuances, they can also complicate the rater's decision-making process and take up valuable rating time. Ultimately, this can lead to a reduction in reliability.

One compromise is to use a trichotomous anchoring system, such as "not done," "partly done," "well done." This approach helps to overcome the tendency of many raters to "give credit" to learners whenever possible while still setting a high standard for performance. It allows organizers to create summary scores that represent the proportion or percent of items rated as "done well" versus "partly or not done." When identifying appropriate behavioral anchors for each of these response options, it is important to consider the level of the learner and the likely distribution of competence in the learner population. Consequently, one can avoid floor (everyone does poorly) and ceiling (everyone does well) effects and maximize the degree to which the items differentiate among trainees. While having a middle option is useful to avoid extreme ratings, it does open up an opportunity for "central tendencies," a non-committal judgment. Thus, there has been a movement toward even-number rating forms which provide a need to determine the quality of the performance while adjusting for very strict and very lenient raters.

Table 2.11 Step-wise process for creating behavior-specific rating form items

1.	**Conceptualize the competencies needed to perform the station task well,** e.g., communication skills, physical exam skills, shared decision-making, delivering bad news
2.	**Compare that conceptualization with available standards,** i.e., literature, experts
3.	**Operationalize the competencies to turn them into written items,** e.g., "uses open-ended questions," "asks about alcohol use"
4.	**Determine the rating options,** i.e., not done/partly done/well done descriptions versus global ratings
5.	**Create behavioral anchors to help evaluators identify which rating option to select,** e.g., well done if done more than once
6.	**Pilot the rating form** multiple times if possible
7.	**Refine the rating form**

Global Rating Form Items

Global ratings address general impressions about a learner's performance in a particular domain (e.g., communication skills, medical knowledge, professionalism), or they may also address overall satisfaction with an encounter. SPs are often asked to indicate the degree to which they would recommend the learner as a clinician to a family member or friend. Similar measures are widely used with "real" patients to assess their satisfaction and perceived quality of care.

Some OSCE organizers are concerned that global ratings are less reliable because they are not anchored in specific, observable behaviors. Thus, they might be more susceptible to rater subjectivity and biases. However, research has not borne out these concerns. Such broad assessments provide an overall "gestalt," and include more intuitive aspects of the

raters' judgments. They may even capture performance elements that are not reflected in checklists. Global ratings are especially important when dealing with advanced learners. For example, professionals with much experience typically use pattern recognition rather than inductive reasoning to arrive at a diagnostic impression. Generally, the reliability of global ratings has been quite satisfactory [14]. Often such ratings also have specific anchors. For example, "Would you recommend this trainee to a family member or friend?" can use simple descriptions such as "not recommend," "recommend with reservations," "recommend," "highly recommend." One can also add more specifics (e.g., respectful, accountable, sensitive to my needs) to better clarify each point on the scale. Since global ratings provide less specific information about what was done well and what needs to be improved, adequate spaces for free-style comments are especially critical when running a formative OSCE.

Best Practices: OSCE Rating Forms

1. Develop rating form items based on the blueprint and ensure that the number of items is optimal for domain coverage, validity and reliability, but not so large that it would overwhelm the rater.
2. Consider using both checklist and global rating items to achieve a balance between identifying specific strengths and areas that need improvement while also providing a more holistic evaluation of the overall performance success.
3. Develop behavioral anchors that consist of observable actions and are likely to maximize performance differentiation among the learner population.
4. Keep the language simple to enhance clarity and leave little to no room for interpretation. Using different font sizes and bolding can help structure the text to help raters complete the evaluations more swiftly.
5. Pilot your rating forms to make sure that all items are well understood and that there is sufficient time to complete the task.
6. If the ratings are provided via an online program, offer adequate training to assure that raters are comfortable with the process and that mistakes and technical glitches can be avoided.

Step 6: Recruit and Train SPs

Recruitment

When choosing SPs, one has to keep in mind that each case has unique requirements—some are physiological, others are psychological. Before initiating the recruitment process, it is important to list all physical or psychological characteristics that would jeopardize the succinct portrayal of a case. Physiological contraindications may include scars, atrophied injection sites of insulin-dependent diabetics, respiratory ailments, heart murmurs, or other physical findings that may diminish the fidelity of the case. Psychological contraindications may include discomfort in exposing one's body if a physical exam is part of that station, or a more aggressive interpersonal style if the case asks for a withholding attitude. If a person is exceptionally outgoing, emotions will reveal themselves non-verbally, which will be more difficult to control in repeated encounters. Even experienced actors can have difficulty overcoming their natural ways of expressing themselves during an extended time period. Casting the right person for the case is important for creating an adequate level of realism. Furthermore, the energy to transfer a very active style into a passive one may distract from other tasks such as remembering history items or evaluating the trainee.

One must also keep in mind that the primary motivation for working with SPs is to create educational programs. Thus, the ability to partner in training or assessment endeavors is another critical characteristic to look for. This can make candidates with teaching experience particularly well suited. In many places, it is not legal to ask about age, race, or body characteristics when hiring. Thus, routinely requesting a realistic photograph can greatly improve casting decisions.

Familiarity with the medical problems can either help or hinder the simulation. On the one hand, having experienced a physical condition oneself may provide special insights into the case. On the other hand, memories about one's own interactions with health care professionals may overshadow the encounter with the learner and may provide a hazard to standardization of the case portrayal or to rater tasks. Furthermore, it can be traumatizing for the SP and lead to early burnout. To avoid an increased need for SP maintenance, it is better to select SPs for whom the medical problem does not evoke special emotions. As Table 2.12 illustrates, some SP characteristics are likely to reduce future training needs.

In general, SPs must be able to control their emotions well. For example, they cannot appear upset if something tragic happened in their real life, and they cannot explode on the learners because they are angry with the project administration. This type of job takes someone who does not burst into laughter if a trainee reacts in an unusual fashion, asks strange questions, or even attempts to make the SP break role. SPs also need to be comfortable in cross-cultural encounters since learners may be from many different backgrounds. In the recruitment phase it can be difficult to detect biases and stereotypes (some are detailed in Step 7) with accuracy, but it is still something to keep in mind when making hiring decisions.

Table 2.12 SP characteristics that simplify training

SP characteristic		Effect on training
Acting experience	→	Less need to train acting (especially of high emotional levels)
Teaching experience	→	More understanding of educational objectives, rating, and feedback tasks
Health care professionals (or trainees in the health professions)	→	More understanding of learner role and technical issues (e.g., interview, physical exam, terminology, culture)
No personal expertise with the case problem	→	Less emotional involvement with the case
Type casted	→	Less need to teach affect
Prior SP experience	→	Less need to teach the mechanics of OSCEs
Over age 18	→	No need for developmental considerations
Less advanced age groups	→	Easier to train, may remember better
GTA,[a] MUTA,[b] PETA,[c] experience	→	Comfortable with physical exams, used to focus on performance details, expert in breast and pelvic or urological exams

There are geographical differences related to potential candidates. Programs in metropolitan areas often have a preponderance of actors, while the SP pool at other locations more typically consists of teachers and a variety of other community members. Some people consider actors to be ideal SP candidates. Whether they are professionals or amateurs, these are people who like to slip in and out of roles and may jump at an opportunity to do so. However, it will be important to clarify for them that working as SP is not a creative act. Even though much improvisation is needed, the focus is on standardization, teaching, and assessment. Not every actor is willing to go along with that, and often the OSCE commitment will be relinquished if a real acting opportunity comes up. To emphasize that SPs are expected to become "educational partners," it is best to refrain from calling them "actors" which can provide the illusion that their main purpose is to give a creative performance.

Over the years, OSCE stations have diversified significantly. In addition to typical encounters with patients and family members, they can also include standardized learners (e.g., in OSTEs) and/or colleagues and team members (e.g., standardized nurses). Stations can be developed to reduce the need for terminology and other professional competencies. However, sometimes it may be helpful to recruit healthcare professionals or their trainees to serve in these SP roles.

To help newcomers to the field develop job descriptions and interview guidelines, the Association for SP Educators (ASPE) developed an online lesson which includes various templates for writing job descriptions and interview questions [15]. They also list a variety of recruitment resources that range from bulletin boards to social media. Once a program has developed a cadre of SPs, word of mouth will often become the most effective and efficient way of recruitment. The advent of virtual and hybrid OSCEs can widen the recruitment net significantly. It is no longer necessary to be in the same geographic location. Remote SPs can be located all over the country or all over the world. This provides new opportunities for developing a diverse SP body which can be necessary to create educational programs that measure up to current inclusivity standards. It can also greatly facilitate last-minute coverage to address sudden SP cancellations.

Over the last decade, organizations such as ASPE have been pushing for more SP professionalization. Moving away from "volunteer" or "independent contractor" status, many institutions started to hire SPs as per diem or part-time employees. Lengthy onboarding procedures can hamper SP recruitment processes significantly, especially in emergency situations, but they can also offer SPs additional benefits and result in a relationship that can be more enduring. For more information, refer to Chap. 12.

Training for Case Portrayal and Delivering Feedback

To make a case come to life, SPs need to become accomplished in three different areas: (1) SPs must know all the physical, psychological, and social details relevant for their case, (2) they must be able to consistently portray the right emotional tone—not too much and not too little, but just the right amount that fits the case, and (3) their actions and responses must be timed correctly. Many novice SPs tend to give away all the information they have about the scenario right up front, maybe even feeling some relief to have gotten the story right. However, often we want learners to practice or demonstrate skills for eliciting information and, thus, sharing this data prematurely eliminates that opportunity. Since OSCE encounters are time limited, it is important that learners have a chance to come to some closure. A continuation of questioning or emotional intensity could make that impossible. For this reason, SPs will need to learn to pace themselves and to adhere to warning knocks or other indicators that the encounter needs to come to an end.

Whenever more than one SP is to be prepared for the same case, group training is necessary for standardization. SPs can read through the case together while clarifications are provided. One can also produce a standard-setting video to emphasize non-verbal behaviors and emotional tone.

Role-playing the case multiple times with trainers as well as each other is essential. It is also helpful to expose SPs to good as well as poor learner performances. By practicing with each other, SPs can gain important insights into the interviewer role and gain empathy for learners.

Table 2.13 lays out a simple protocol for training SPs for case portrayal. There are varied opinions as to how much training is necessary for SPs to perform their case adequately. A classic book on SP training advocates a 5-session approach: (1) Familiarization with the Case, (2) Learning to Use the Checklist, (3) Putting it All Together (Performance, Checklist, Feedback), (4) First Dress Rehearsal, and (5) Final Dress Rehearsal [16]. The total amount of training time will depend on OSCE and case requirements, cost, and time limitations. Gliva-McConvey and colleagues created a "Human Simulation Continuum Model" which demonstrates the increase in training intensity based on the level of standardization needed [17].

If the OSCE is formative, two hours may be adequate, especially with experienced SPs. If it is a summative assessment, training will have to be much more extensive (there have been literature reports of 10–20 hours of training, divided into training segments). Typically, 2 hours is the limit to how much SPs can absorb at one time. Higher stake events are likely to require a trial run. With experienced SPs who know the locale, online training can be sufficient even for in-person OSCEs. However, when organizing remote OSCE encounters, it is crucial that the training sessions use the same modality to ensure log-in access, Wi-Fi stability, and appropriate background.

The ASPE Standards of Best Practice provide a variety of guidelines that range from training (e.g., reviewing station objectives) to reflecting on the training process (e.g., quality management) [18].

Best Practices: SP Recruitment and Training

1. Search for SPs through word-of-mouth strategies (e.g., by contacting other SPs, connecting with other SP trainers).
2. Cast the right person for each case (i.e., physical appearance, psychological profile, availability, no contraindications).
3. For high-stakes programs, recruit and train alternates who can step in if needed (alternates can be cross-trained to provide coverage for multiple cases).
4. Put SPs into learner's positions through roleplay to enhance their understanding of the case (e.g., interactive and emotional impact of SP actions) and to promote an empathic approach to learners.
5. Practice all aspects of the encounter (e.g., physical exam, feedback); do not leave SP performance to chance.
6. Explore the psychological and physiological impact a case has on the SP to avoid toxic side effects (e.g., getting depressed from repeatedly portraying a depressed patient, getting muscle spasms from portraying a patient who has difficulty walking).
7. Train all SPs who are portraying the same case (simultaneously or consecutively) together to enhance standardization.
8. For multi-site OSCEs with different sets of SPs, it is important to create detailed training protocols and standard setting videos to enhance comparability.

Table 2.13 Training protocol for SP case portrayal (if multiple SPs portray the same case, they need to be trained together to promote standardization)

Explain	• Training program purpose and structure, project history • Logistics (location, date of program(s), etc.) • Who they will work with, who else will be in the room • Resources to stop psychologically or physically unsafe encounters/situations
Explore	• SP expectations and concerns • Relationship to case in personal life or SP work • Past experiences with learners at the targeted or different training level—how did it go?
Review	• Read all station materials aloud (e.g., objectives, rating forms, feedback instructions). • SP instructions should be read by SP(s) in 1st person, stop along the way to explain, elicit emotional reactions, jointly come up with additional information to round out the case (e.g., name of spouse). Clarify: – Case content, story, what information needs to be conveyed – Emotional tone, type, and intensity – Timing of SP interventions: what to say/do in the beginning, middle, end of the encounter and things revealed only upon prompting by the learner – Clarify 2-3 standardized phrases that should occur in all encounters
Practice	• Role-play multiple times with trainer and other SPs portraying the same case • Different interview styles and topics that are not already included in the SP instructions • Put SPs in the role of the learner to explore their perspective and understand the level of case difficulty
Pilot[a]	• Organize trial runs under real-life OSCE conditions (location, timing, learners similar to targeted group)

[a]If exam leads to significant promotion decisions, or complex OSCE structure

9. Emphasize the importance of stopping an encounter or getting help if the simulation or a learner creates an unsafe situation.

Step 7: Recruit and Train Evaluators

When planning an OSCE, it is important to determine who will rate the participants' performance. On the surface, faculty raters may appear ideal, but they are not necessarily more accurate [19] and often have limited availability. Many programs use SP raters since they can achieve a good level of reliability, offer the "patient" perspective, are more easily trained, and their availability is already established. Self- and peer assessments as well as observing SP ratings can also be considered. At times, evaluations are completed by more than one type of observers (e.g., SPs and observing faculty) focusing on pre-assigned components of the assessment. This is the ideal situation because it considers different perspectives and results in a more comprehensive, accurate, and fair assessment.

Regardless of who completes the rating, attention must be given to raters to provide as accurate and reliable ratings as possible. The rater task is difficult because there are so many factors that can interfere with precise performance assessment. Generally, there are three elements to rating a learner's performance: (1) observation of specific behaviors (technique and content), (2) judgment of the behavior against a set of standards, and (3) documentation of the rating. Problems can occur during each of these rater tasks as illustrated in the rater self-assessment guide in Table 2.14. There has been some debate whether rating skills are innate or whether they can be taught. A combination of selection and training are likely to yield the best results.

Some people are naturally more aware of their biases and are inclined to look for evidence more objectively rather than to follow their lenient or stringent tendencies. If organizers have the luxury to select raters, an initial screening could consist of assembling candidates in small groups and showing them selected videos of station encounters. By setting a required level of inter-rater and test-retest reliability, one can quantify the suitability and readiness of the candidates in question. In projects where major promotion decisions depend on OSCE performance, one may even go as far as certifying observers.

Rater training works best if it occurs in a group setting. Table 2.14 details a series of strategies that will help make raters aware of the types of errors they are more likely to make and their personal rating style, whether they are "doves" (i.e., easy raters) or "hawks" (i.e., harsh raters). Self-awareness is no guarantee of being completely error free, but it is the best chance to provide a fair rating. Some institutions have developed calibration videos so that raters can practice. After scoring a sample performance (often online), raters compare their results with those of experts to improve objective ratings. However, group calibration training that includes discussion is likely to be more effective.

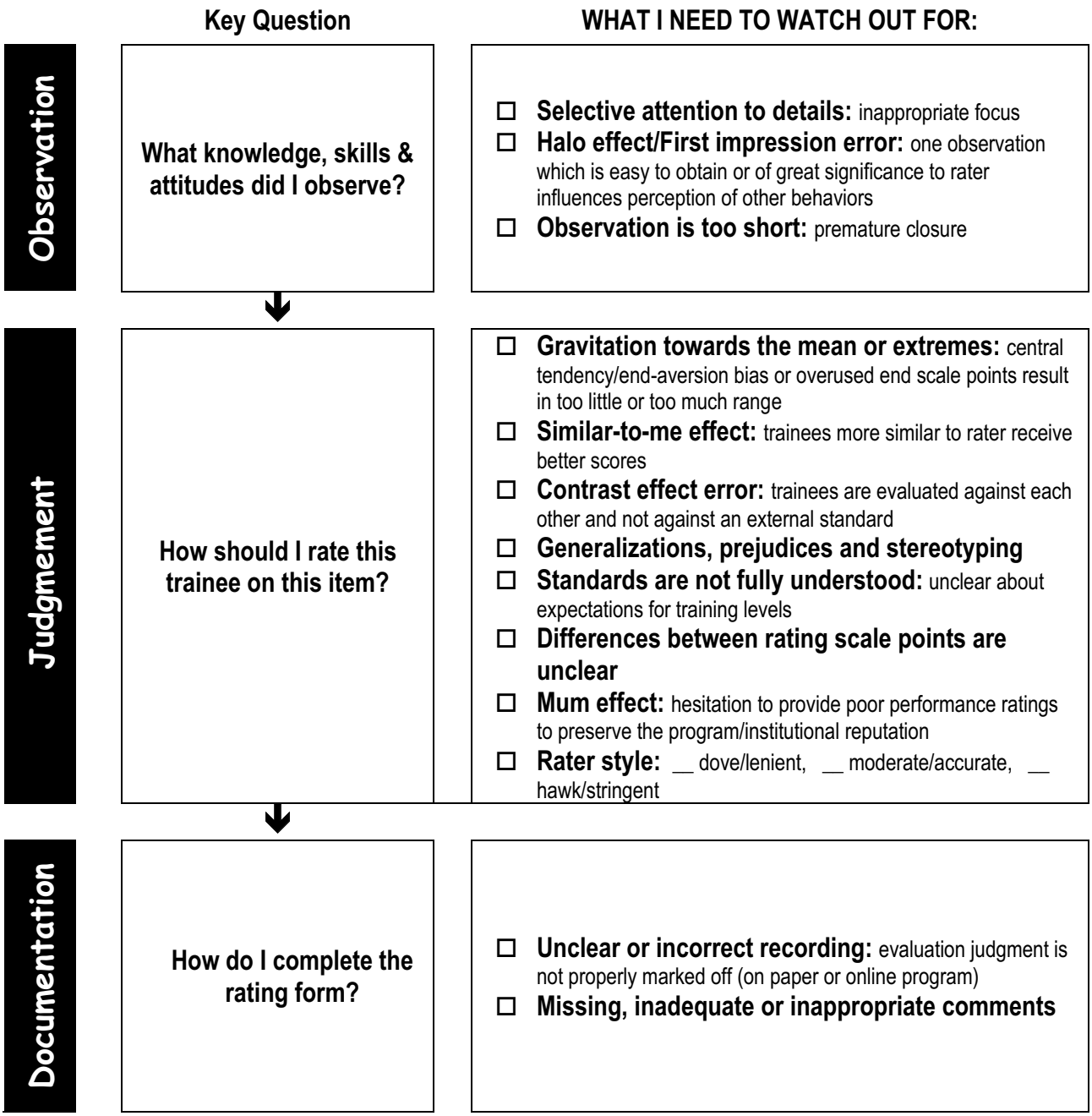
A rater training protocol is detailed in Table 2.15. The amount of training time will vary significantly depending on who the raters are, how much rating and OSCE experience they already have, how stringent the assessment is and how much time is available. With clinician raters, it may be most difficult to carve out some training time if no compensation can be provided. However, they also need some type of training to orient them to the goals, process and content of the exercise. At a minimum, this must be done in writing. Providing raters with feedback about their past ratings (e.g., tendencies, similarities/differences with other raters) will also go a long way to help evaluators understand what they do well and what they could improve on.

Attitudes and emotions undoubtedly play a central role in the rating process. It is important for trainers to be aware of how raters feel about the project and their task. Since not everybody can be involved in OSCE assessment development, raters must at least understand the underlying rationale and feel confident that categories were not selected arbitrarily. Rater trainers must continuously encourage questions. Although questions add to training time, they are better dealt with before the OSCE starts than while it is in progress or when the project is over and one realizes that a rating form item has been completely misunderstood.

In formative OSCEs, raters are often also asked to provide feedback in verbal and/or written form. Typically, there are time/space limitations and feedback providers need to be brief but specific enough for learners to take subsequent actions. Over the years, a variety of feedback models have been developed. Table 2.16 provides a sample protocol that could help structure a brief verbal feedback session. By adding 2-3 station-specific teaching points, one can also ensure that the teaching objectives for each station are accomplished with each learner. Providing succinct and meaningful feedback is not always easy. If raters are also expected to give feedback to the learner, they should practice doing so in advance of the OSCE.

In addition to individualized in-station feedback, many formative OSCEs also include post-OSCE group debriefings. Typically, they incorporate all the learners who just completed all the stations and often they also incorporate faculty and/or SPs. This is yet another opportunity to discuss what worked and what did not work in each of the stations. A major benefit of the large group debrief is that it promotes peer teaching and role modeling.

Table 2.14 Rater self-assessment guide to help improve accuracy



Best Practices: Evaluator Recruitment and Training for Rating and Feedback Tasks

1. Select evaluators who are willing to adopt the program values, who are consistent in their ratings and don't have their own agenda.
2. Bring multiple evaluators together to jointly observe a learner performance on tape or live, compare ratings, and discuss similarities and discrepancies. Practice giving feedback (if this is expected).
3. Make raters aware of potential biases and rating mistakes.
4. Provide written guidelines for rating items, evaluation scheme, and station objectives/teaching points.
5. Post-OSCE, give feedback to raters about how their ratings compare with those of others (e.g., more or less lenient, lack of range).

Table 2.15 Rater Training Protocol

1.	**Provide a training program overview:** e.g., • Training program purpose and structure • Logistics (location, date of program(s), etc.) • Who they will work with, how much time will be available, what are the program objectives, project history, what a typical encounter looks like, who else will be in the room • Resources to stop psychologically or physically unsafe encounters/situations
2.	**Explore rater expectations/concerns** • Past experiences with learners at the targeted or different training level—how did it go; how did it compare to expectations? • How might this rating experience be similar/different from previous rater work? • Rater expectations and concerns
3.	**Review the station(s) to be observed and rated** • Provide a copy of the rating form and define each item (including examples for the response options) • Provide all other case materials (including learner and SP instructions)
4.	**Review typical rater errors** • Discuss factors that can interfere with rating tasks (see self-assessment form in Table 2.14), encourage raters to become aware of their own style and tendencies
5.	**Perform multiple practice ratings** • Use live encounters or videos to demonstrate a "gold standard" evaluation and to establish intra- and inter-rater reliability • Compare ratings within the group until a consensus is reached • Help raters pace themselves by using OSCE-specific time frames (if possible, organize trial runs in the place where the OSCE will take place)
6.	**Introduce raters and SPs** (if rating is done by a faculty or SP observer) • Explain how rater and SP evaluations are different yet will complement each other. Typically, faculty raters evaluate global entrustment and medical knowledge while SPs often focus on communication and professionalism • Encourage raters and SPs to work together without sharing their individual impressions about the learner's performance before documenting their own ratings • Request that the SPs play through the case with the raters taking on the learner role to build understanding of the case challenges and build appropriate empathy for the learners

Table 2.16 Feedback training protocol

1.	**Provide a feedback framework/model** • Explain the behavior change model which helps diagnose learners as pre-contemplative, contemplative, ready for action, in maintenance or relapse stage. Using this framework, feedback can be tailored to optimize its impact on learning. • Share learner feedback about the feedback (i.e., what learners gained from feedback in post-OSCE debriefing sessions or on program evaluation forms)
2.	**Introduce characteristics of effective feedback—written or verbal** • Create a psychologically safe environment to reduce defensiveness • Specific not general • Focus on behaviors that can be changed, not on personality or other unchangeable characteristics • Take advantage of all observers in the station (if applicable) • Connect station content with previous experiences (e.g., have you had a similar case?) • Explore what could be done differently next time (feed forward)
3.	**Provide feedback anchors** (i.e., teaching points specific for the case that should be covered to strengthen the overall message/station objectives)
4.	**Practice giving feedback** (e.g., utilizing video or role play, preferably illustrating good and poor performances)

Step 8: Implement the OSCE: Managing the Session

In addition to station-specific materials (Table 2.8), it is also necessary to develop forms and other resources that help with the overall organization of the event. Table 2.17 details the various forms that will be needed. Table 2.18 provides an example station rotation schedule for all OSCE participants (numbers 1-8 refer to the case number, "rest" refers to the rest station). By identifying a station number, and circling the numbers across the table, one can follow the faculty or SP's sequence. Such a form is also included as a blank worksheet in the back of this book (Appendix J) along with blank program evaluation surveys (Appendix K: Participant program evaluation survey; Appendix L: Rater program evaluation survey; Appendix M: SP program evaluation survey).

Whenever one plans an event that involves a large number of people, organization can be challenging. One must accept the fact that the unexpected will arise. However, with good planning and adequate resources, one should be able to make the program successful. To make troubleshooting at the time of the OSCE easier, it is helpful to contemplate potential solutions ahead of the event. Key concerns include attendance, standardization, timing, and emotion management. Organizers should ask themselves what they could do in the event of the contingencies listed in Table 2.19. We have included some solutions that have worked for us. By having extra SPs and faculty on hand, one can overcome lateness,

Table 2.17 General OSCE materials (in addition to station-specific materials in Table 2.8)

Forms	Purpose	Content elements	Considerations/tips
Faculty orientation and debrief guide	To standardize OSCE experience for all learners	• Description of timing and types of learning activities • Key learning points for each case	• Consider meeting with faculty to prepare them with guide
Learner orientation materials	To record attendance and assign ID codes (if applicable) To pre-brief learners and increase psychological safety	• OSCE name/number, location, date • Learner names and ID codes • Explain OSCE process and basic assumptions	• If multiple sites or OSCE implementations, consider a script or brief orientation video for consistency • Provide consent forms (if appropriate)
SP/Rater orientation materials	To record attendance and match SP/rater names with ID codes (if applicable)	• OSCE name/number, location, date • SP/rater names and ID codes	• Permit room for multiple SPs per station if alternates • Allow room for comments and to record unexpected occurrences (e.g., lateness) • Provide forms for SPs or others to receive payments
Rotation Schedules (Example: Table 2.17)	To guide the flow of the OSCE, indicate what station learners start with and track where they should be at any given time	• OSCE name/number, location, date • List of participant names/IDs • Areas for indicating expected and actual rotation time periods • Station sequence • Rest stations or general breaks (if applicable)	• Add time parameters as reminder (e.g., minutes allowed for SP encounter) • Allow room for comments and to record unexpected occurrences • Provide room for monitor(s) name(s) • For virtual or hybrid OSCEs, indicate links and contact information for SPs, raters, learners and tech help to assure a reliable and swift backchannel • For virtual encounters, add pictures of learners to reduce rotation errors (important for SPs and raters who are unfamiliar with the learners)
Learner resource guide	To provide additional resources for learning aligned with OSCE cases	• Could include links, graphics, evidence-based articles or guidelines	• Can be provided electronically for easy reference after the OSCE • Resource guide can be updated annually by faculty
Learner Post-OSCE program evaluation forms (Example: Appendix K)	To evaluate the OSCE experience	• Prior exposure to clinical tasks/cases • Realism of stations • Evaluated my performance fairly • Provided a good learning opportunity	• Keep it brief • Comments can provide interesting qualitative data • Anonymity can increase response rate and validity
SP and Faculty Program evaluation forms (Example: Appendix M & L)	To evaluate the OSCE	• Level of case difficulty • Educational value for learners • Faculty development value (if applicable) • SP performance (if faculty rating) • Appropriateness of case • Effectiveness of instructions/preparation	• Comments can provide interesting qualitative data • Consider questions related to blueprint validation (e.g., did the station address the station objectives)

early departures, or absences. Adequate training, extra props, and forms can help with standardization and with enhancing the fidelity of the simulation. Timing and SP/faculty adjustments may be necessary to keep the OSCE on schedule. Organizers and monitors need to be on the lookout for nervous learners who may enter stations too early or tired SPs who do not portray their case correctly anymore. Post OSCE debriefing is essential: this includes SPs, faculty, and everyone else involved in the simulation program. For SPs, it is especially important to de-role from cases that may be emotionally challenging to portray repeatedly.

Best Practices: Optimizing the OSCE Environment

1. Familiarize yourself with the OSCE location (layout, bathrooms, SP changing room) and its resources (e.g., printing options, station assignment options).
2. The OSCE schedule should take into account adequate time for setup and arrival as well as pre-briefing and debriefing learners, SPs, and faculty.
3. The number of times SPs can accurately perform their tasks will vary by case scenario, rating, and feedback

responsibilities. Adequate breaks or switch outs may be essential to assure quality work.
4. If the location is not a simulation center, testing staff should include one proctor for each hallway and one overall administrator.
5. Online and hybrid OSCEs need to have stable internet connections and sufficient backup plans. Adequate staffing and robust communication strategies also need to be in place to address potential problems.
6. Conduct a "dress rehearsal" prior to any high-stakes OSCE.

Table 2.18 Sample OSCE rotation schedule

	Date: *March 19th (PM session)* *Brief orientation from 12:45–1pm* *Each learner has 8 encounters and 2 rest stops* *Last OSCE encounter ends at 4:00pm. Faculty and residents debrief from 4-4:30PM.*									
	Timing									
Learner name	*1:00*	*1:18*	*1:36*	*1:54*	*2:12*	*2:30*	*2:48*	*3:06*	*3:24*	*3:42*
Dr. A	1	2	3	4	Rest	5	6	7	8	Rest
Dr. B	Rest	1	2	3	4	Rest	5	6	7	8
Dr. C	8	Rest	1	2	3	4	Rest	5	6	7
Dr. D	7	8	Rest	1	2	3	4	Rest	5	6
Dr. E	6	7	8	Rest	1	2	3	4	Rest	5
Dr. F	5	6	7	8	Rest	1	2	3	4	Rest
Dr. G	Rest	5	6	7	8	Rest	1	2	3	4
Dr. H	4	Rest	5	6	7	8	Rest	1	2	3
Dr. I	3	4	Rest	5	6	7	8	Rest	1	2
Dr. J	2	3	4	Rest	5	6	7	8	Rest	1

Table 2.19 OSCE troubleshooting: potential problems and possible remedies

What if…
☐ **Someone doesn't show?** For high-stakes OSCEs, always cast extra SPs. Cross-training for multiple stations can cut down costs. Extra faculty who are familiar with multiple stations can also fill in at the last minute if needed. For formative OSCEs, ask a faculty member to portray the patient or have learners rotate through stations in pairs if multiple circuits are operating at the same time.
☐ **Someone has to leave temporarily?** Participants and SPs should be informed in advance when designated breaks will occur. For long exams, breaks need to be scheduled in and it is a good idea to cast and train multiple SPs for individual stations. While this requires more extensive training to standardize performance and rating across SPs, it ensures an "understudy" will always be on hand.
☐ **A rater does not complete the forms correctly on paper or online?** Keep rating forms and the submission process as simple as possible. Designate a staff member to regularly review completed forms during the OSCE so rating errors can be corrected as quickly as possible. Virtual and hybrid OSCEs need to have robust backchannels that allow for instant communications among all parties involved.
☐ **A participant enters the wrong station?** Make sure exam proctors are monitoring the exam and can make timely substitutions in the rotation schedule if needed. Online stations (aka breakout rooms) also need to be well-labeled and monitored.
☐ **Timing is off-schedule?** If a station goes over the allotted time, try shortening subsequent rotations by small increments until the schedule is back on track. Communicate adjustments to SPs and faculty in the stations to keep them fully informed about last-minute administrative decisions.
☐ **Someone is late or has to leave early?** Add adequate arrival time before the OSCE to mitigate potential transportation problems. Prepare for potential delays by adding some extra time at the end of the OSCE to premature departures by SPs or faculty. Maintain open communications with SPs and faculty to learn about potential problems as they arise.
☐ **An SP does not portray the case correctly?** Schedule ample training so that everyone is satisfied with the case portrayal before the actual OSCE. Make sure there is a staff member familiar with all cases present to answer any specific questions that may arise in student encounters. Try to catch mistakes as early as possible (e.g., real time observation by SP alternates or trainers) to intervene quickly.
☐ **Station materials are missing?** Bring extras of everything, including any props and all forms. Determine in advance the easiest way to make emergency paper copies. For online or hybrid O SCEs, organize electronic materials for easy distribution.
☐ **Some stations consistently take less than the allotted time?** Check in with the SP between rotations; adjust details of the case portrayal if needed. With formative OSCEs, use the extra time for extra feedback. Avoid early dismissal from the station. Learners who congregate in hallways or online breakout rooms often discuss station content which can prejudice them in subsequent encounters.

Step 9: Manage, Analyze, and Report Data

It is important to identify resources and make a plan for entering, managing, and analyzing data early on in the OSCE development process so that you do not end up with poor quality or un-interpretable data—or worse yet, missing data. To do this, "begin with the end in mind" by clarifying what information you hope to obtain from the OSCE and plan accordingly. We have found that good data management practice—which includes protecting trainee privacy through the use of unique IDs and secure storage—is crucial because it not only ensures high-quality data but also helps create a safe learning environment for your trainees. How you handle, use, and report trainee data may be dictated by institutional policy, accreditation regulations, or the law. If you anticipate wanting to conduct research using OSCE data, it is particularly important to understand local policies and regulations with regard to treating trainees as human subjects (e.g., Institutional Review Boards) early in the planning process, see Chap. 7 "Scholarship and Education Research Registry" for further details.

Managing Data

Since it is likely that multiple people will be involved in handling the data from an OSCE, good data management principles should be employed to ensure confidentiality and the integrity and security of the data. Tables 2.20 provide a step-by-step approach to addressing privacy concerns.

Ideally, data from OSCEs should be entered directly by raters into user-friendly computer interfaces that then download the data into formats that can be readily uploaded into statistical analysis and data visualization software (e.g., SPSS, R, SAS) for analysis. If paper rating forms are used, it is good practice to collate data as soon as possible in order to be able to identify any problems with the quality of the data (e.g., inconsistent ratings, missing data, missing learner IDs) and to be able to resolve any problems while memories are still fresh (e.g., data are missing because someone arrived late). This is true of digital and online data collection systems as well—reviewing the data and the video recordings (when included within a simulation software system) soon after the OSCE is an essential quality assurance practice in order to identify and ameliorate issues when one still can.

While data can be initially entered into a spreadsheet (e.g., Microsoft Excel or Google Sheets), which is familiar to most people, we recommend the use of structured data entry forms that facilitate fast, consistent, and less error-prone data recording. Data collected within structured forms or software platforms are easily exported into a wide range of analyzable formats while ensuring that data fields are accurately labeled. Such forms can be created in "off the shelf" software (e.g., Microsoft Access) or using "open-source" free programs (e.g., RedCAP [20], Epi Info [21], Formsite [22]). Structured data entry forms use specific fields for each assessment data point, determining which options can be selected, providing a label for those options (e.g., 0 = No, 1 = Yes or 0 = Not Done, 1 = Partly Done, 2 = Well Done), and also a label for the item itself. Thus, field-based data entry also facilitates the creation of a "data dictionary" that provides all of this information. In addition, data dictionaries can include essential "meta-data" about the OSCE (sessions and dates, location, trainees, any issues or problems that may have occurred), station/case names and details, and raters (often a good practice to develop an ID system for identifying SPs and/or other assessors so that names are not stored with the data). It is always a good idea to have an OSCE summary sheet that lists important details about each OSCE: date, location, learners, raters, cases, problems, where data is stored and status of data, etc. Refer to Chap. 6, "Data for Learning and Program Evaluation: Analyzing, Reporting and Visualizing OSCE Data," for more information on data processes and pipelines that maximize the uses of OSCE data.

Analyzing OSCE Data

Start with descriptive statistics such as distributions of ratings across the response categories (frequencies) for each item on the checklist to identify data entry errors and missing data. Once the quality of the data has been assured (and descriptive statistics rerun if necessary), summarize the data across learners to identify program-level gaps in train-

Table 2.20 Creating a secure and confidential OSCE data storage system

1.	Generate a unique ID for each individual learner.
2.	Link the unique ID to each learner's identifying information (name, email, class/program, other system IDs) in a file (database or spreadsheet.)
3.	Store that file securely by password-protecting the file and saving on a secure server (encrypted institutional server) or within a secure institutional system. Do not save to personal or shared devices or to personal cloud storage devices.
4.	Limit access to the linking file to only those who require identified data (e.g., those with responsibility for learner assessment).
5.	Store the OSCE data without identifiers, using only the unique ID (delete names, emails, and any other identifying information).
6.	Ensure data back-up systems are in place and monitor that regular back-up is occurring.

ing for specific skills and to establish norms for the group or check against predetermined standards of expected performance (see Fig. 2.2). Reviewing the data in this detail will help in understanding how to summarize the data for individual learners and for the cohort of learners and will also provide guidance to improve assessments.

Calculating and Interpreting OSCE Scores

Feedback on trainee performance can be shared at the level of the individual assessment items but often "scores" are calculated that reflect ratings across multiple items, cases/stations, and/or the entire OSCE. The reasons for calculating OSCE scores are to derive summaries of performance that can then be used to (1) set minimum standards for high stakes, pass/fail examinations; (2) provide feedback to learners (and their faculty) on performance at a more generalizable level (i.e., a competency across clinical scenarios); and (3) provide overall feedback to your program on the effectiveness of training.

Scores can be based on averages of scaled items (items with response options that represent a numeric scale) or on percentages; the latter are used especially for checklist scores (e.g., % of behaviors rated as "done" or "well done"). If scaled items are non-normally distributed because response options represent a ranking but no clear numerical interpreta-

Fig. 2.2 Distribution of OSCE checklist item rating for the Annual General Internal Medicine OSCE (n = 20 residents across 8 cases)

Communication Skills	Not Done	Partly Done	Well Done
Information Gathering			
Started with **open-ended** questions	Started with closed, yes-no questions **5%**	Began with open-ended questions but stopped prematurely **60%**	Started with open-ended questions and continued using them as appropriate **35%**
Asked **what you thought** was the matter	Did not specifically ask **35%**	Asked but did not give you enough time to share views **25%**	Asked so that you fully shared your views **40%**
Managed the **narrative flow** of your story	Not able to elicit your story because questions not organized logically **10%**	Elicited main elements of story, but illogical order of questions disrupted flow **30%**	Elicited full story by asking questions that facilitated natural flow of story **60%**
Elicited your story using **appropriate questions**	Impeded story by asking leading questions or more than one question at a time **8%**	Used leading questions and/or asked more than one question at a time, but still able to share most of story **38%**	Facilitated the telling of your story by asking questions one at a time without leading you in your responses **55%**
Clarified information by repeating to make sure he/she understood you on an ongoing basis	Did not clarify (did not repeat info you needed) **20%**	Repeated the information but didn't give you a chance to indicate whether accurate **30%**	Repeated information and directly invited you to indicate whether accurate **50%**
Allowed you to talk **without interrupting**	Interrupted you **10%**	Did not interrupt you directly but cut your response short by not giving you enough time **20%**	Did not interrupt and allowed time to express thoughts fully **70%**
Relationship Development			
Communicated **concern** or intention to help	Did not communicate intention to help/concern via words or actions **5%**	Words OR actions conveyed intention to help/concern **20%**	Actions AND words conveyed intention to help/concern **75%**
Non-verbal behavior enriched communication (e.g., eye contact, posture)	Non-verbal behavior was negative or interfered with communication **15%**	Non-verbal behavior demonstrated attentiveness **20%**	Non-verbal behavior facilitated effective communication **75%**
Acknowledged your emotions appropriately	Did not acknowledge your emotions **15%**	Attempted to acknowledge your emotions **40%**	Responded to your emotions in ways that made you feel better **45%**
Was **accepting/nonjudgmental**	Expressed judgement **0%**	Did not express judgement but did not demonstrate respect either **20%**	Demonstrated respect towards you **80%**
Used words you understood and/or explained **jargon**	Jargon made it difficult to understand **5%**	Used jargon occasionally but did not significantly interfere with understanding **25%**	Provided no opportunity for misunderstanding by avoiding or explaining jargon **70%**

tion (e.g., poor, fair, good, excellent), non-parametric statistics can be used (e.g., Cochran's Q, Friedman's Chi Square, Wilcoxon Signed Ranks). For each OSCE, multiple scores can be calculated:

1. Overall OSCE Scores: For each station, calculate a summary score (e.g., % of maximum points achieved, mean of scaled items). Then, summarize across stations by computing an average (of summary scores for each case) or summing up the station scores across the OSCE. It is best to calculate station scores only when the station was designed to assess a defined skill-set as an overall score (e.g., physical examination, history gathering, communication, etc.). In an OSCE station calling for performance of many skills, an overall score can obscure relevant information because it creates one summary score that collapses differences across multiple skill domains.
2. Domain Scores: For each station, calculate sub-scores (e.g., % of maximum points achieved, mean of scaled items, % of items "done well") for the items representing specific domains or categories of skill/performance (e.g., communication skills, counseling, physical examination). Then average or sum up the sub-scores across all stations where a particular domain was assessed.

When designing a blueprint (Step 3), one needs to make sure that each competency/domain is assessed in more than one station. Thus, learners have more than one opportunity to demonstrate their skills, and the degree to which the context may influence performance (i.e., case-specificity) can be determined. More samples provide a more reliable indication of competence. Although there is no hard and fast rule, generally, specific skills should be assessed across a minimum of three cases in order to achieve minimum reliability. In most OSCEs, the same core communication skills are assessed in every case because interpersonal and communication skills typically generalize across clinical scenarios. Consequently, most assessments report "communication" performance as a summative (across-cases) score.

Assessing the Quality of the OSCE Data

Whenever one organizes an assessment of competence for summative purposes, one needs to be concerned with a variety of psychometric standards, focused mainly on establishing the reliability and validity of the measure. Table 2.21 provides definitions of these key psychometric concepts, describes the questions they address, and provides information on strategies for enhancing the quality of the assessment.

When evaluating the quality of your OSCE data, the first question to explore is: to what degree do ratings of learners' performance across the OSCE stations consistently assess learners' underlying competence? This question focuses on inter-station reliability or the internal consistency of the items which assess specific domains across items and stations and are then used to derive summary OSCE scores. Estimates of internal consistency, or the degree to which sets of assessment items "hang together" (i.e., that a learner who does well on such items in one case will do well on those items in another case), can be calculated using Cronbach's alpha (available in most statistical software programs). Calculating Cronbach's alpha can also identify problematic items—items that were not used consistently by SPs, that were worded in ways that interfered with interpretations, or that do not end up reflecting performance in a particular station. Deleting these items may improve the overall internal consistency of items compromising a summary OSCE score. In most statistical software programs, output for Cronbach's alpha can include what the alpha would be for each set of items if that item were deleted, showing whether individual items enhance or attenuate overall reliability. Cronbach's alphas range from 0 to 1 and generally, estimates above 0.80 suggest that items are internally consistent. For pilot OSCEs (newly developed and undergoing initial evaluation) and OSCEs with fewer stations (less than five or six), Cronbach's alphas should probably exceed 0.60 or 0.70. These types of OSCEs focus on providing learners with practice and formative feedback and are not used to make summative judgments about competence. The consistency of the checklist can also be assessed by estimating test-retest reliability (comparing performance scores for trainees who complete the same OSCE or case at two different times, without intervening training) and inter- or intra-rater reliability (comparing checklist ratings among different raters or over time within the same rater).

Once the reliability of an assessment (e.g., a checklist) has been established, attention should turn to gathering evidence of its validity: the degree to which it measures what it was intended to measure. The first step in evaluating the validity of an assessment involves scrutinizing the content of the assessment, through expert and literature review, to ensure that it appears to capture or reflect the chosen competence or practice. After this initial screen, there is no single or simple way to establish validity and instead efforts to support the validity of a checklist should be based on how well it performs over time and across situations—establishing validity is based on an accumulation of evidence. Validity evidence comes from answers to the following kinds of questions: Does the assessment discriminate among trainees at different levels? Is performance in the OSCE, as measured by the checklists, significantly associated with other measures of related skills (e.g., patient satisfaction, faculty, and peer ratings)? And ultimately, are checklist scores predictive of actual clinical performance and outcomes?

Table 2.21 Psychometric qualities of OSCE results

	Definition, key questions	Enhancement strategies
Reliability & internal consistency	• Measures consistency and precision of an assessment tool. If learners underwent the same exam without any interim interventions, would the results be the same? How similar did trainees perform in the different stations? Typically, one uses Cronbach's alpha to determine the level of internal consistency (a Cronbach's alpha between .60 and .80 is considered adequate for formative assessments, an alpha of .80 or more is necessary for promotion decisions). • Typical sources for unreliability are: ☐ Item differences within cases (case specificity) ☐ Case differences in the use of the rating form ☐ Differences within individual raters in how they applied the rating form ☐ Differences between raters in how they applied the rating form	Sufficiently large sample size: ☐ Of learners ☐ Of cases (e.g., samples of communication abilities) ☐ Clear, easy-to-use rating forms ☐ Training of raters ☐ Strong evidence of test item importance ☐ Elimination of items that are responsible for reducing the OSCEs reliability
Intra-rater reliability	• Measures consistency of individual raters over time. If a rater would evaluate the same performance a second time, would the result be the same? Contextual differences (e.g., live versus video-taped encounter or multiple viewings of the same video-taped encounter), are expected to influence these estimates of reliability. Nonetheless, if the rating forms are reliable, we would expect to see substantial correlations.	☐ Initial selection of raters who are consistent ☐ Rater training (including feedback on the correlations of assessments of the same video-taped case at different times)
Inter-rater reliability	• Measures consistency among different raters. If several raters observe the same learner's performance, are their ratings of the performance in agreement?	☐ Initial selection of raters who are consistent ☐ Rater training (including feedback on the level of agreement with other raters of the same real or video-taped encounter)
Validity	• Determines whether an OSCE assesses what it is set out to measure (e.g., communication skills, primary care skills). There are multiple types of validity. – Face and content validity (Does it look right?) – Convergent/divergent validity (Does it compare to other measures as it should?) – Discriminant validity (Does it differentiate between training levels or other learner characteristics) – Predictive validity (Does it predict future behavior)	☐ Re-examination of the blueprint

Standard-Setting

Setting standards for pass/fail examinations is both an art and a science. The core issues are determining the appropriate developmental level, and then exploring how to use score cutoffs to divide learners into those that meet these standards and those that do not. For high-stakes examinations, many psychometricians and medical education experts recommend absolute or criterion-referenced cutoffs (i.e., scores that reflect the ability to competently perform specific skills and behaviors). Experts review the "test" (OSCE cases and rating forms) content and determine a "passing" score. More complicated methods are also available [23–25]. Refer to Chap. 6, "Data for Learning and Program Evaluation: Analyzing, Reporting and Visualizing OSCE Data," for more information on standard-setting.

An alternative is to use relative standards or norm-referenced standards, where a certain percentage of the lowest-performing OSCE participants "fail" (e.g., those with a score in the bottom decile or the bottom 20%). The obvious problem with this approach is that while the pass/fail cut-off often stays the same, the sample of OSCE participants may vary in their performance (e.g., a score in the bottom decile in a class of stellar students might be comparable to an average score in a class with greater variation in their skills). This approach also requires that at least some trainees "fail."

Standard setting policy decisions are judgments made by experts. Formal standard setting procedures can assist in ensuring that cut-off scores reflect a consensus among relevant responsible educators. A variety of standard-setting processes have been described for performance-based assessments, each with its own underlying assumptions and requirements [26]. While exams given on a very large scale can afford—both financially and with respect to having sufficient numbers of subjects and experts—to go through rigorous standard setting procedures, most smaller-scale projects cannot. Therefore, many school or program-based summa-

tive OSCEs end up using an approach that combines normative, criterion-based, and practical considerations to setting pass/fail cutoffs.

At NYU Grossman School of Medicine, we use this combined approach for setting cutoffs to identify students who fail our comprehensive clinical skills exam (CCSE), a summative 8-station OSCE required after the core clerkship year. Through rigorous training of raters and refinement of our checklists and patient note rating processes, we are able to obtain internally consistent assessments of the four competence areas assessed in the exam: communication skills, history gathering, physical exam, and clinical reasoning (the latter is reflected in the patient note). These scores are normally distributed around a mean score between 50% and 60% and therefore we can identify students at both the upper and lower ends of the spectrum. We have decided that performing well on one competency does not compensate for performing poorly on another. Therefore, we report the competency scores (calculated across all cases) separately, taking what is called a non-compensatory approach [27]. We then set a normative passing cutoff at the lowest decile for each competency. Students in this lowest decile across 2 or more competencies and students who "fail" communication skills alone (because we have found that this is predictive of failure on the USMLE Step II CS exam) are identified and added to this list. Students' scores that fall close to the threshold (above and below) are further scrutinized to better make pass/fail decisions. Finally, any student who received a "would not recommend to a friend or family" global rating from more than one SP is added to the list, because we have found this identifies additional students who go on to struggle with communication issues clinically and on other OSCEs. Our list of students who fail the exam is based also on our capacity to provide adequate remediation. Remediation strategies are discussed further in Chap. 3, "Remediation of Learners Who Perform Poorly on an OSCE."

Reporting Results

If the OSCE is used solely for training, performance feedback is essential. Even if the OSCE has evaluative purposes, students want and highly value feedback on their performance. Because of the need to keep the content of the OSCE stations secure, there may be limitations on how detailed such reports can be. Training program faculty need to know how learners performed. By identifying those areas of consistent weakness across learners, the curriculum can be modified to enhance learners' clinical performance in the future. Figure 2.3 provides an example of an OSCE score report with communication scores across all cases (% well done across all communication items as well as within specific communication sub-domains like relationship development and information gathering). Whether in the form of a table or with the help of graphs, learners need to know what scores they achieved and how they compared with their peers. Learners can be encouraged not only to compare their scores with those received by peers but also to explore their relative strengths and weaknesses, noting differences among how they performed within and across particular domains. We aspire to design feedback reports to be easily understood and build in opportunities to develop action plans and ongoing guidance to learners as part of the feedback process.

Longitudinal Educational Database

OSCEs generate a wealth of data and can be combined with data from other sources (e.g., faculty ratings, exam scores, self-assessments, even clinical and patient data) over time to track, monitor and understand the development of competence. You can work with your local Institutional Review Board to develop opportunities for obtaining consent from learners to combine those data not just for program evaluation purposes but also for research purposes—to answer both anticipated and unanticipated questions about the longitudinal process of becoming competent professionals. A student or trainee "registry" can be established, just like a patient registry, in which all students or trainees are asked to provide permission for their routinely collected educational data to be linked and compiled in an educational database. Such data, once linked and stored, should be de-identified—all identifiers should be stripped from the data except for the unique ID generated for the purpose of the database. Creation of this database can provide invaluable data on performance across domains over time and also help establish the quality of assessments made throughout the curriculum. For further details about research registries, please see Chap. 7, "Scholarship and Education Research Registry."

Best Practices: Managing and Analyzing OSCE Data

1. Plan for and monitor the quality of data entry and management; use unique identifiers to maintain confidentiality and make sure data are backed up and maintained securely.
2. Explore the quality of the data in terms of reliability estimates of internal consistency (Cronbach's alpha) before calculating summary scores.
3. Calculate OSCE scores based on performance within domains across stations, considering the structure of the data (response options) and how best to derive summaries

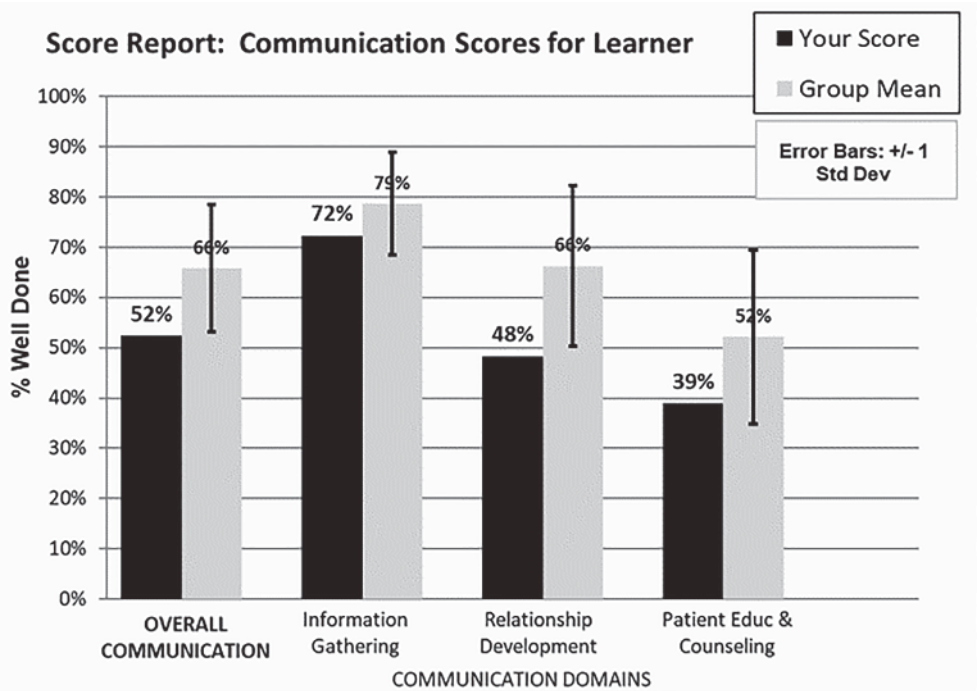


Fig. 2.3 Sample report card illustrating the OSCE performance of an individual learner (following our example General Internal Medicine Residency OSCE). Scores are reported as % checklist items "well done" and reflect individual performance across 8 OSCE cases relative to their cohort of 20 1st year resident OSCE participants

(% well done, average of scaled items, non-parametric methods if necessary).

4. Report performance data to learners in ways that are understandable and constructive.
5. Consider how to mine the wealth of educational data available by creating registries and organizing and linking data and information from many relevant sources.

Step 10: Develop a Case Library and Institutionalize OSCEs

The first OSCE requires an especially great deal of effort. However, a collection of cases is created, station materials are developed, a cadre of SPs is recruited, and the team involved gets more experience and, therefore, organizing OSCEs becomes much easier. By developing a case library such as the one exemplified in Table 2.22, one can greatly reduce preparations for subsequent OSCEs. It is useful to maintain a library in electronic and paper format. Since repeated usage typically results in revisions, one also needs to make sure that the latest versions of the cases are archived. Deciding on a consistent naming, dating, and labeling strategy for the revisions early on will be critical for using the database efficiently. Adding learner performance data and program evaluation results (e.g., from faculty, SPs or learners) that may have accumulated for each station will also prove useful in future OSCE plannings.

It is also helpful to maintain an SP Database and keep track of the stations they worked on previously. The learning management systems of some simulation centers will track SP utilization. Additional notes about how well they portrayed the cases and performed their rating and feedback tasks will help determine who should be invited back to work on the same or similar stations. In addition to contact information, it is essential to add a realistic photograph which will be important for matching SPs to new stations.

OSCEs can involve many different data sets which can be pulled together in a relational database. Automatic linkage between the Station Bank, the SP Database, Performance and Program Evaluation data will greatly facilitate future OSCE planning. Even without such software expertise, it is helpful to keep in mind that the integration of data available for each station will assist with subsequent planning as well as potential scholarship efforts.

OSCEs are a mainstay of health professions education since they are a valuable tool for training as well as for evaluating simple and complex skills. It makes sense for organizers to invest energy in institutionalizing OSCEs and incorporating them in the institutional culture.

Table 2.22 Example case bank

Case			SP characteristics		Nature of encounter			Skills assessed				
ID	Name	Details	Age	Gender	Acute v Chronic	Presenting Condition	Visit Setting	General	Case Specific			
								Communication	History Gathering	Physical Exam	Treatment & Management	Other
1	Chest Pain	Patient with history of NIDDM and HTN presents with chest pain	56	F	Acute	NIDDM, HTN	ER	✓	✓	✓	✓	Interprofessional collaboration
2	Depression	Precept MS IV on ambulatory rotation	SL = 25	M	Chronic	Case presented: 64 y/o pt with depression	Outpatient follow- up	✓	✓		✓	Precepting
3	New Diagnosis	Patient hospitalized while on vacation	55	M	Acute	High BP, Afib	Outpatient follow- up	✓	✓	✓	✓	System-based practice
54	Diarrhea	Patient complaining of diarrhea	35	M	Chronic	Diarrhea	Virtual - Telehealth	✓	✓		✓	Telehealth skills
55	Shoulder Pain	Patient diagnosed with RCC 8 months ago	62	F	Acute	Metastatic RCC	Urgent Care	✓	✓	✓	✓	
56	Stomach Ache	Patient complaining of stomach pain; worse with GAD	32	F	Chronic	GAD	follow- up	✓	✓	✓	✓	
57	Cholesterol	Patient with history of HTN and hyperlipidemia, noncompliant	45	M	Chronic	HTN, hyperlipidemia	Outpatient new visit	✓	✓		✓	Up to date guidelines
58	Work Rounds	Sub-intern and team conduct work rounds	SL = 27	F	Acute	Case presented: 58 y/o pt with MBC	Inpatient Rounds	✓			✓	Precepting

Best Practices: Building Institutional Capacity

1. Save all material on an institutional server.
2. Create a collaborative interdisciplinary OSCE committee that meets regularly. Share resources with other departments.
3. Invite institutional opinion leaders and early adaptors to observe and/or take on valuable roles in station creation and OSCE implementation.
4. Disseminate reports widely and demonstrate how helpful your OSCEs are regarding accreditation, program evaluation, and teacher and learner satisfaction.
5. Integrate the OSCE into your departmental budget.
6. Promote the OSCEs throughout the year.

References

Step 1

1. The Society for Simulation in Healthcare (SSH): https://www.ssih.org
2. The Association for SP Educators (ASPE): https://www.aspeducators.org

Step 2

3. Poenaru D, Morales D, Richards A, O'Connor HM. Running an objective structured clinical examination on a shoestring budget. Am J Surg. 1997;173(6):538–41. https://doi.org/10.1016/s0002-9610(97)00007-x.
4. Reznick RK, Smee S, Baumber JS, et al. Guidelines for estimating the real cost of an objective structured clinical examination. Acad Med. 1993;68(7):513–7. https://doi.org/10.1097/00001888-199307000-00001.
5. Brown C, Ross S, Cleland J, Walsh K. Money makes the (medical assessment) world go round: The cost of components of a summative final year Objective Structured Clinical Examination (OSCE). Med Teach. 2015;37(7):653–9. https://doi.org/10.3109/0142159X.2015.1033389.

Step 3

6. MedEdPORTAL: https://www.mededportal.org/

Step 4

7. ASPE Case Development Template: https://www.aspeducators.org/aspe-case-development-template
8. Lewis KL, Bohnert CA, Gammon WL, et al. The Association of Standardized Patient Educators (ASPE) Standards of Best Practice (SOBP). Adv Simul (Lond). 2017a;2:10. https://doi.org/10.1186/s41077-017-0043-4.
9. Jain V, Agrawal V, Biswas S. Use of formative assessment as an educational tool. J Ayub Med Coll Abbottabad. 2012;24(3–4):68–70.

Step 5

10. Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33(3):206–14. https://doi.org/10.3109/0142159X.2011.551559.
11. Kachur EK, Green S, Dennis C. Written comments on objective structured clinical examination rating forms: an exploratory study. Teach Learn Med. 1990;2(4):225–31. https://doi.org/10.1080/10401339009539465.
12. Vu NV, Marcy MM, Colliver JA, Verhulst SJ, Travis TA, Barrows HS. Standardized (simulated) patients' accuracy in recording clinical performance check-list items. Med Educ. 1992;26(2):99–104. https://doi.org/10.1111/j.1365-2923.1992.tb00133.x.
13. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. Acad Med. 1999;74(10):1129–34. https://doi.org/10.1097/00001888-199910000-00017.
14. Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. Med Educ. 2003;37(11):1012–6. https://doi.org/10.1046/j.1365-2923.2003.01674.x.

Step 6

15. The Association for SP Educators (ASPE): https://www.aspeducators.org
16. Wallace P. Coaching standardized patients: for use in the assessment of clinical competence. Springer Publishing Company; 2007.
17. Gliva-McConvey G, Shannon GM, Pitt J, Clark L. The human simulation continuum: integration and application. In: Gliva-McConvey G, Nicholas CF, Clark L, editors. Comprehensive healthcare simulation: implementing best practices in standardized patient methodology. Cham: Springer; 2020. https://doi.org/10.1007/978-3-030-43826-5_5.
18. Lewis KL, Bohnert CA, Gammon WL, et al. The Association of Standardized Patient Educators (ASPE) Standards of Best Practice (SOBP). Adv Simul (Lond). 2017b;2:10. https://doi.org/10.1186/s41077-017-0043-4.

Step 7

19. Kalet A, Earp JA, Kowlowitz V. How well do faculty evaluate the interviewing skills of medical students? J Gen Intern Med. 1992;7(5):499–505. https://doi.org/10.1007/BF02599452.

Step 9

20. RedCAP: https://projectredcap.org

21. Epi Info: https://www.cdc.gov/epiinfo
22. FormSite: https://www.formsite.com
23. Boulet JR, Murray D, Kras J, Woodhouse J, McAllister J, Ziv A. Reliability and validity of a simulation-based acute care skills assessment for medical students and residents. Anesthesiology. 2003;99(6):1270–80. https://doi.org/10.1097/00000542-200312000-00007.
24. Kilminster S, Roberts T. Standard setting for OSCEs: trial of borderline approach. Adv Health Sci Educ Theory Pract. 2004;9(3):201–9. https://doi.org/10.1023/B:AHSE.0000038208.06099.9a.
25. Kramer A, Muijtjens A, Jansen K, Düsman H, Tan L, van der Vleuten C. Comparison of a rational and an empirical standard setting procedure for an OSCE. Objective structured clinical examinations [published correction appears in Med Educ. 2003; 37(6):574]. Med Educ. 2003;37(2):132–9. https://doi.org/10.1046/j.1365-2923.2003.01429.x.
26. Downing SM, Haladyna TM, editors. Handbook of test development. Lawrence Erlbaum Associates Publishers; 2006.
27. Sadler DR. Interpretations of criteria-based assessment and grading in Higher Education. Assess Eval High Edu. 2005;30(2):175–94. https://doi.org/10.1080/0260293042000264262.

Remediation of Learners Who Perform Poorly on an OSCE

3

Adina Kalet, Linda Tewksbury, Lynn Buckvar-Keltz, Barbara Porter, Calvin L. Chou, and Rachel H. Ellaway

Data from well-designed OSCEs help educators identify trainees with gaps in their core clinical skills. A consensus on effective remediation strategies for individuals who perform poorly is developing [1–3]. Experts stress that it is important to clearly delineate the implications and consequences of learner failure in any performance assessment [4, 5] and assert that successful remediation requires approaches tailored to identified deficits [6]. Effective remediation should be guided by data valid for this purpose (see Chap. 2, Step 9, for an in-depth discussion of standard setting). By definition, learner remediation then must have a reasonable chance of leading to an improvement in clinical competence [7]. Table 3.1 breaks down the remediation process into manageable steps. Also crucial, to engage meaningfully in and gain life-long benefit from remediation, learners must have or develop the capacity to accurately self-assess and self-regulate learning.

Table 3.1 Steps in the remediation process

1.	**Gather and carefully review objective data of performance**
2.	**Obtain student self-assessment and provide feedback based on objective data**
3.	**Assess for nonacademic issues**
4.	**Make an educational diagnosis**
5.	**Formulate an individualized learning plan with diagnosis specific remediation strategies** (Think creatively about available resources!)
6.	**Make a plan to follow-up on progress**

Initiating the Remediation Process

Not surprisingly, trainees are usually very upset upon hearing they have failed an OSCE. A structured first meeting between the student and faculty member responsible for remediation, which allows enough time for discussion of feelings, a student's self-assessment, and a careful review of data from the exam, is reassuring to the student and will most likely to lead to an effective remediation. Using an evidence-based coaching model, such as the facilitated feedback model described by Sargeant et al. [8], is recommended. Depending on the nature of the OSCE (low stakes/formative vs. high stakes/summative), the remediation process can be more or less comprehensive. For a low-stakes exam, a brief individual feedback session, with videotape review if available, may be sufficient. Table 3.2 provides an outline for a more comprehensive intake meeting in a high-stakes situation. We schedule 1.5 h for this initial session.

Using detailed data from the OSCE in remediation is invaluable because it addresses learner resistance to the process, builds accurate self-assessment skills, and provides support for promotion decisions. These data may include the various sources of information listed in Table 3.3.

A. Kalet (✉)
Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA
e-mail: akalet@mcw.edu

L. Tewksbury
Department of Pediatrics, New York University Grossman School of Medicine, New York, NY, USA

L. Buckvar-Keltz · B. Porter
Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA

C. L. Chou
Department of Medicine, University of California San Francisco School of Medicine, San Francisco, CA, USA

R. H. Ellaway
Department of Community Health Sciences, University of Calgary Cumming School of Medicine, Calgary, AB, Canada

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_3

Table 3.2 OSCE remediation: initial diagnostic interview

□ **Statement of expectations**
□ **Learner self-assessment**
□ **Assessment of exam-specific performance issues**
□ **Educational history** Including screening for verbal and nonverbal learning disabilities, attention deficit disorders, and language fluency
□ **Assessment of professionalism** e.g., learner attitudes toward the OSCE and accountability for performance
□ **Screening for situational stressors**
□ **Screening for common psychiatric illness** e.g., depression, anxiety, bipolar disorder, eating disorders, and substance use disorder

Table 3.3 Learner data useful for remediation

□ **Performance across OSCE cases compared to the group means**
□ **Performance by case**
□ **SP patient notes**
□ **SP comments** (after prescreening)
□ **Videotape of the encounter**
□ **Other evaluation data available** e.g., academic record and clerkship comments

Table 3.4 Areas of difficulty leading to poor OSCE performance

1.	**Pre-existing academic issues** • Learning disabilities • Poor academic track record especially on stressful clinical rotations • Nontraditional educational paths such as learners with discontinuous training (e.g., MD-PhD programs) or transfer from other programs (e.g., accelerated BS-MD programs)
2.	**Isolated clinical skills deficit** i.e., specific area(s) of weakness such as knowledge base, communication, reasoning, or problem-solving skills
3.	**Metacognitive or specific testing issues** • Time management or organizational difficulties • Insufficient preparation or poor understanding of the exam • Performance anxiety
4.	**Extenuating psychological factors** • Anxiety • Depression • Situation-specific duress
5	**Nonverbal learning disorders** e.g., long-standing social awkwardness and autism spectrum disorders
6.	**Professionalism issues** i.e., learner does not know or agree with health profession ethical tenets and values; paranoid, combative, or defiant personality style or frank personality disorder

Characterizing the Difficulty

There are a host of reasons learners fail an OSCE. The most common reasons for failure are summarized in Table 3.4 in order of frequency, based on our experience with clinical skills remediation for 3 consecutive classes of medical students, 2007–2009, during which 53 of 500 students failed. For a more detailed table, see Kalet and Chou [9]. Once the faculty facilitator and the learner come to a negotiated agreement on one or more areas of difficulty, a contract or individualized remediation plan (IRP) should be drafted and follow-up plans made. This document (see Fig. 3.1 for an example) should evolve as the remediation process proceeds and new light is shed on the student's strengths and weaknesses. Keeping the IRP updated provides an efficient communication tool among the members of the remediation team and keeps the student actively engaged in the process.

In Table 3.5, we list remediation strategies we use regularly. Relevant references include Dohms et al. [10], Kalet and Chou [2], and Pinsky and Wipf [11] for videotape review; Bowen [12], Croskerry [13], and Mutnick and Barone [14] for clinical reasoning and critical thinking; and Kogan et al. [15] and Puscas et al. [16] for direct observation with feedback. The primary purpose of any strategy is to enhance the learner's awareness of deficits, enabling them to improve their clinical performance and therefore the quality and safety of the care patients' receive. Strategies used will depend on what the learner is struggling with, available resources, and the learner's willingness to explore difficult issues.

Fig. 3.1 Individualized remediation plan example

	Learning Goals	Strategies	Time Frame
1.	Improve my rapport building skills, especially non-verbal expressions of attention and concern	Practice with an SP until I can do this consistently (2–3 sessions).	Perform adequately on Remediation OSCE.
2.	Improve my clinical reasoning so that I can include more pertinent negative historical and physical exam facts in a patient note.	Write up 10 practice cases and review these with Dr. X. Respond to feedback by demonstrating commitment to learning.	Perform above the threshold on Remediation OSCE.
3.	Explore my attitudes about patients who are seeking pain medication.	Write and discuss 3 brief essays: Physicians' attitudes toward pain management in terminally ill patients; Barriers to adequate pain relief in chronic pain syndromes; Ethical issues when treating pain.	By the Remediation OSCE Dr. X will be satisfied that I have explored my own attitudes and beliefs in this area.

Table 3.5 Selected remediation strategies

1.	**Self-directed videotape review (VTR)** Using a blank OSCE checklist, the learner rates his/her performance on one or two videotaped encounters from the actual OSCE, preferably cases where the student underperformed, summarizes his/her findings from the VTR, and reviews these documents with a faculty adviser.
2.	**Faculty-facilitated videotape review** In learners who demonstrate poor self-awareness of their difficulties, a structured, faculty-facilitated VTR can help the student recognize areas of difficulty.
3.	**Peer-group videotape review** Learners who agree to work together, using a framework of mutual support, can provide each other with feedback and encouragement as well as opportunities for corrective experiences through roleplaying alternative approaches to the patient.
4.	**SP practice with feedback** A learner with very specific communication difficulties can benefit from scheduled sessions with an SP experienced in giving feedback, to practice their skills.
5.	**Clinical reasoning practice** Learners are assigned reading about the clinical reasoning/critical thinking process to enhance metacognitive awareness and then practice with paper or web-based cases.
6.	**Direct observation with real patients** This can be done in real time or using VTR with patient permission.
7.	**Physical exam workshops** This can be done in groups with a faculty or resident facilitator; active practice and discussion about how the findings contribute to clinical decision-making is critical to success.
8.	**Reflective writing** Learners reflect on attitudes and beliefs expressed or demonstrated and how they may align or not align with medical professionalism, medical professional identity formation, or effective patient care.
9.	**Directed readings** These are relevant when there is an isolated knowledge deficit or lack of understanding of specific principles such as the tenets of medical professionalism, institutional policies, or standards of treatment (e.g., substance use disorder).
10.	**Work with a specialist** e.g., referrals for learning or psychiatric assessment, interpersonal skills coaching, performance anxiety strategies, learning/organizational support, and career advice.

Who Should Participate in Learner Remediation?

The most effective facilitators of clinical competence remediation are likely to be, but not restricted to, experienced clinician educators. Table 3.6 lists examples of the experts and specialists who we have found are invaluable to the effort depending on the specific needs of the struggling student.

Table 3.6 Experts and specialists who can contribute to learner remediation

1.	**Clinical educators** Best suited to conduct the initial assessment, work with learners on clinical reasoning or physical examination deficits, monitor remediation process, and make a final outcome determination.
2.	**Communication skills coach** Learners with isolated communication deficits or professionalism issues benefit from working with a coach familiar with the healthcare environment and skilled with behaviorally focused coaching approaches.
3.	**Drama therapist/SP trainer/experienced SP** Learners with communication skills deficits or performance anxiety benefit from practice with feedback and coaching.
4.	**Learning specialist/studying or executive function coach** Learners with a long-standing history of uneven academic performance, atypical organizational and unusual study strategies or those who don't "read for pleasure" may have undiagnosed learning disabilities, valuable to diagnose because either targeted strategies and accommodations are available or at the least a trainee becomes more self-aware and learns to effectively protect patients and advocate for their own needs.
5	**Psychiatrist/psychologist** Best when a psychiatric diagnosis is suspected or already established or trainee requires emotional support as a consequence of being identified for remediation.
6.	**Role model** A respected member of the clinical field related to the learner's interests can be effective at encouraging the student to engage in the remediation enthusiastically.

Table 3.7 Learning objectives for clinical skills remediation faculty development

Clinical educators conducting remediation with learners who fail an OSCE should be able to…	
1.	Interpret quantitative and qualitative data regarding the competence of individual medical trainees.
2.	Define clinical competence in a behaviorally specific, measurable manner.
3.	List common areas of difficulty for trainees struggling to pass an OSCE.
4.	Discuss models of normal adult development which influence clinical competence development.
5.	Describe learning theories which are salient to performance-based assessment and remediation (e.g., mastery learning, competency-based medical education, sociocultural learning theory, transformative learning theory).
6.	Describe the screening process needed to identify a learning disability or attention deficit disorder.
7.	Demonstrate the ability to screen for common psychiatric issues that may manifest as or coexist with clinical incompetence.
8.	Make defendable judgments regarding clinical competence.
9.	Effectively coach student through the remediation process, promoting a growth mindset.
10.	Document a remediation process that is meaningful and addresses legal and regulatory requirements.
11.	Explore personal attitudes, beliefs, and experiences that may impact a struggling learner's ability to engage in the remediation process, with attention to one's own personal biases.
12.	Understand that on rare occasions a student may fail the make-up exam and need guidance to navigate the consequences including career planning.

Faculty Development for Remediation

The institutional capacity to remediate learners who struggle with a high-stakes OSCE is entirely dependent on the number, commitment, and expertise of the faculty members available to participate. Faculty members who are good listeners, skillful at giving effective feedback, familiar with effective learning skills, knowledgeable about individual characteristics associated with common learning challenges and psychological barriers to performance assessment [17], and interested in the development of clinical competence and working with learners one-on-one are ideally suited for this work. These faculty may still need additional training to maximize their effectiveness, and faculty development focused on coaching skills that promote a growth mindset can be particularly helpful [18]. It is also critical for the educator to be attentive to the diverse personal experiences, attitudes, and beliefs that may impact the ability of struggling learners to engage in the remediation process while being attentive to one's own biases [19]. Table 3.7 lists learning objectives for faculty development in clinical skills remediation. Educators specifically interested in reading more about defining behavioral measures of clinical competence are referred to Quirk [20].

Make-Up OSCE

A remediation program, to be effective, must culminate in a measure of learner performance. In remediation for high-stakes exams, we have required students to participate in and pass a four-station OSCE, which is a mix of cases repeated from the OSCE they failed and new cases. Because reliability of a four-station OSCE is predictably poorer than one with more cases, we determine the outcome of this exam using standards established in the larger exam and take into account findings from a detailed review of the student's performance. Each case is videotaped or directly observed by a faculty familiar with the student.

Considerations When Documenting Remediation

Detailed documentation of the remediation process is important both to ensure communication among the remediation team and to provide evidence to support promotion decisions. At the minimum, programs should keep track of learner's data on OSCEs, standards for pass/fail decisions, and individualized remediation plans and document the date and time of meetings between learners and members of the remediation team. We have found it helpful to write a brief narrative summary of each session with a learner, documenting updates to the individualized remediation plan and agreed upon next steps. Depending on the local law and regulatory environment, schools and training programs have obligations and responsibilities to keep written records of the evidence that learners have demonstrated training-stage appropriate competence. Remediation team leaders should familiarize themselves with the professional accreditation body (ACGME, LCME) and institutional documentation requirements that may apply to the remediation process. In the United States, in addition to documentation requirements for the purposes of accreditation, there is relevant federal law that seeks to protect the privacy of students, patients, and employees by limiting access to records. The Family Educational Rights and Privacy Act of 1974 (FERPA [21]) applies to a student's educational records, and the Health Insurance Portability and Accountability Act of 1996 (HIPAA [22] applies to the student's personal health. Balancing the needs to document a complex process in a meaningful way and understanding the legal environment will help each program design an efficient record keeping process, which serves both the program and the learners. On rare occasions, a remediation is unsuccessful. The institution must be able to accept this outcome and justify the consequences, all while supporting the student with empathy and career advice [23, 24].

Remediation of learners who perform poorly on an OSCE provides a unique opportunity to explore the underlying reason(s) for substandard clinical skills and to ensure impactful intervention. Although some of these learners are challenging to engage and motivate, most of the time, a remediation experience is a rich opportunity for professional and personal growth in the student as well as development of a therapeutic alliance between the learner and remediation specialist(s). In our experience, most students gain valuable insight regarding their difficulties, are committed to working with the remediation team, and successfully complete the make-up academic exercises. Work remains to be done regarding the identification of the most effective, efficient, and least costly remediation techniques for the various sub-types of problems leading to failure on clinical skills examinations.

References

1. Chou CL, Kalet A, Costa MJ, Cleland J, Winston K. Guidelines: the dos, don'ts and don't knows of remediation in medical education. Perspect Med Educ. 2019;8(6):322–38. https://doi.org/10.1007/s40037-019-00544-5
2. Kalet A, Chou CL. Remediation in medical education: a mid-course correction. 2nd ed. Springer; 2023. https://doi.org/10.1007/978-3-031-32404-8
3. Kalet A, Chou CL, Ellaway RH. To fail is human: remediating remediation in medical education. Perspect Med Educ. 2017;6(6):418–24. https://doi.org/10.1007/s40037-017-0385-6
4. Cleland J, Arnold R, Chesser A. Failing finals is often a surprise for the student but not the teacher: identifying difficulties and supporting students with academic difficulties. Med Teach. 2005;27(6):504–8. https://doi.org/10.1080/01421590500156269
5. Sayer M, Chaput De Saintonge M, Evans D, Wood D. Support for students with academic difficulties. Med Educ. 2002;36(7):643–50. https://doi.org/10.1046/j.1365-2923.2002.01259.x
6. Hauer KE, Ciccone A, Henzel TR, et al. Remediation of the deficiencies of physicians across the continuum from medical school to practice: a thematic review of the literature. Acad Med. 2009;84(12):1822–32. https://doi.org/10.1097/ACM.0b013e3181bf3170
7. Ellaway RH, Chou CL, Kalet A. Situating remediation: accommodating success and failure in medical education systems. Acad Med. 2018;93(3):391–8. https://doi.org/10.1097/ACM.0000000000001855
8. Sargeant J, Armson H, Driessen E, Holmboe E, Könings K, Lockyer J, Lynn L, Mann K, Ross K, Silver I, Soklaridis S, Warren A, Zetkulic M, Boudreau M, Shearer C. Evidence-informed facilitated feedback: the R2C2 feedback model. MedEdPORTAL. 2016;12:10387. https://doi.org/10.15766/mep_2374-8265.10387
9. Kalet A, Chou CL. Remediation in medical education: a midcourse correction. 1st ed. Springer. Chapter 2: An Example of a Remediation Program; 2013. p. 17–37. https://doi.org/10.1007/978-1-4614-9025-8
10. Dohms MC, Collares CF, Tibério IC. Video-based feedback using real consultations for a formative assessment in communication skills. BMC Med Educ. 2020;20(1):1–9. https://doi.org/10.1186/s12909-020-1955-6
11. Pinsky LE, Wipf JE. A picture is worth a thousand words: practical use of videotape in teaching. J Gen Intern Med. 2000;15(11):805–10. https://doi.org/10.1046/j.1525-1497.2000.05129.x
12. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. New England Journal of Medicine. 2006;355(21):2217–25.
13. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med. 2003;78(8):775–80. https://doi.org/10.1097/00001888-200308000-00003
14. Mutnick A, Barone M. Assessing and remediating clinical reasoning. In: Kalet A, Chou CL, editors. . Remediation in medical education: a mid-course correction; 2014. p. 85–101. https://doi.org/10.1007/978-1-4614-9025-8
15. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. JAMA. 2009;302(12):1316–26. https://doi.org/10.1001/jama.2009.1365
16. Puscas L, Kogan JR, Holmboe ES. Assessing interpersonal and communication skills. J Grad Med Educ. 2021;13(2 Suppl):91–5. https://doi.org/10.4300/JGME-D-20-00883.1
17. Kua J, Lim WS, Teo W, Edwards RA. A scoping review of adaptive expertise in education. Med Teach. 2021;43(3):347–55. https://doi.org/10.1080/0142159X.2020.1851020

18. Deiorio NM, Moore M, Santen SA, Gazelle G, Dalrymple JL, Hammoud M. Coaching models, theories, and structures: an overview for teaching faculty in the emergency department and educators in the offices. AEM Educ Train. 2022;6(5):e10801. https://doi.org/10.1002/aet2.10801
19. Bright CM, Duefield CA, Stone VE. Perceived barriers and biases in the medical education experience by gender and race. J Natl Med Assoc. 1998;90(11):681–8. PMID: 9828583 PMCID: PMC2608378
20. Quirk M. Intuition and metacognition in medical education: keys to developing expertise. Springer; 2006.
21. Family Educational Rights and Privacy Act (FERPA). U.S. Department of Education Website. https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html. Published 1974.
22. The Health Insurance Portability and Accountability Act of 1996 (HIPAA). U.S. Department of Education Website. https://aspe.hhs.gov/reports/health-insurance-portability-accountability-act-1996. Published 1996.
23. Stringham RVV, Whitlock J, Perez NA, Borges NJ, Levine RE. A snapshot of current US medical school off-ramp programs—a way to leave medical school with another degree. Med Sci Educ. 2021;31(2):341–3. https://doi.org/10.1007/s40670-020-01175-w
24. Petersen KH, Jain NR, Case B, Jain S, Solomon SL, Meeks LM. Compassionate off-ramps: the availability of terminal master's degrees in US medical schools. J Med Educat Curri Develop. 2023;10:23821205231164022. https://doi.org/10.1177/23821205231164022

Beyond the OSCE: Using Standardized Patients in the Clinical Setting

4

Sondra Zabar, Harriet Fisher, Jeffrey Wilhite, Zoe Phillips, Renee Heller, Colleen Gillespie, and Kathleen Hanley

Standardized patients (SPs) are widely used in assessment; however, the majority of SP exercises involve "announced" encounters in which the learner knows that they are interacting with a simulated patient. While Objective Structured Clinical Examinations (OSCEs) are considered a gold standard for assessing clinical skills, their contrived nature may limit their ability to capture the true behaviors of medical professionals [1]. Unannounced standardized patients (USPs) are a relatively new, but increasingly employed, method of evaluating the skills of medical professionals [2, 3]. Unlike OSCEs, USP encounters are in situ assessments in which the learner is not aware that the patient is an SP. As such, there are no artificial time constraints nor potential for observation bias, and the USP can evaluate the in situ behaviors of clinicians and the clinical system.

USPs have been used internationally to assess trainees and practitioners across health professions, including nursing [4, 5], optometry [6, 7], and a variety of medical specialties, from primary care [8, 9] to emergency medicine [10, 11]. USPs are employed to assess a range of items, including clinical skills [12, 13] and the efficacy of educational interventions [14, 15]. Clinicians' performance when conducting a visit with a USP can also be compared to their performance in announced standardized patient (ASP) encounter in practice or an OSCE [16, 17]. Other studies explore the feasibility and validity of USPs compared to the corresponding medical chart and clinical vignettes [18]. Starting in 2019, USPs have been sent to pharmacies to assess medication dispensing practices [19, 20].

USPs simulate the entire experience of a real patient before, during, and after a clinical visit, which allows them to provide comprehensive feedback on care quality from the moment they are first connected to the care team to the time they complete their post-visit assessment. Unlike SPs in OSCEs, who are only able to assess the clinician during the encounter, USPs can be used to evaluate the clinical microsystem (i.e., clerical staff, medical assistants, and other members of the care team patients may encounter during their visit) as part of the larger, learning health system. The standard model for primary care, the "patient-centered medical home," prioritizes the integration of the clinical microsystem in patient care and is as important to health outcomes as the skills of physicians [21]. The clinical care team's ability to collaborate and communicate well can be easily assessed from the USP's perspective, as can the overall experience of navigating the clinic [11, 18]. Further, USPs can be employed to ensure adherence to national patient safety standards (such as hand washing and patient identification), evaluate clinical preparedness for emerging infectious diseases [22], and observe a clinic's level of patient centeredness [23].

In addition to assessing the larger system, USPs can be deployed to assess significant components of the clinical experience, including communication related to social determinants of health (i.e., housing instability) and subsequent referral to social services [24], screening practices to identify underlying depression or risk behaviors [9, 25], safe medication prescription practices [25–27], and healthcare costs [28]. USPs are capable of evaluating the entire range of communication skills employed during a routine medical visit which include core areas such as information gathering, relationship development, and patient education and counseling, as well as a patient's more subjective experiences that are consequences of those communication skills (i.e., satisfaction and degree of post-visit activation) [29]. USPs provide clinical leadership with a comprehensive

S. Zabar (✉) · H. Fisher · J. Wilhite · Z. Phillips · R. Heller
K. Hanley
Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA
e-mail: sondra.zabar@nyulangone.org

C. Gillespie
Institute for Innovations in Medical Education, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_4

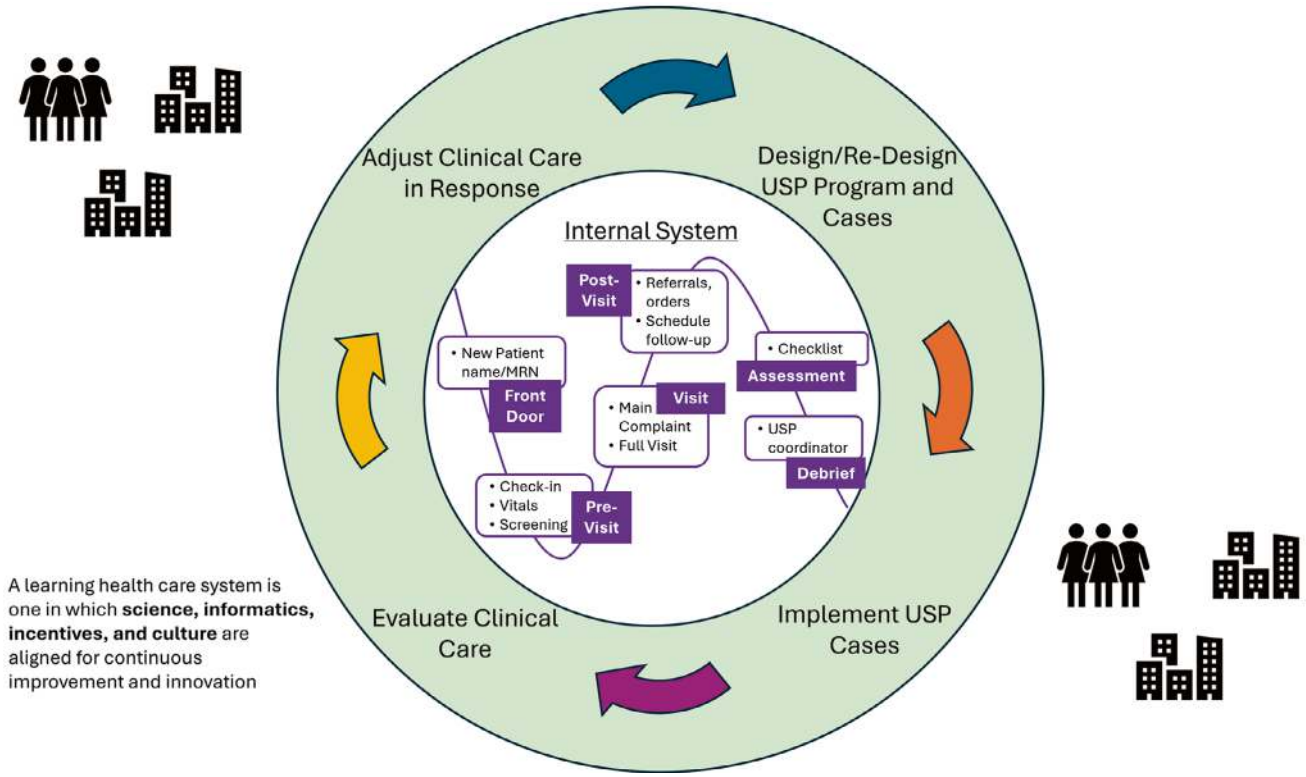


Fig. 4.1 The clinical microsystem

view of on-the-ground clinician practices and a patient's view of the clinical microsystem.

By offering objective and consistent feedback, USPs contribute to evidence-based assessment and continuous refinement of the learning healthcare system (Fig. 4.1 [35]). USPs have been widely deployed to systematically identify performance deficits in communication, but may be easily adapted for quality improvement within and between institutions [30]. Though single USP visits provide only a snapshot of clinician care, deployment of USPs over months or years—particularly with regular opportunities for feedback and intervention—can identify variation in performance and inform recommendations for system-level change and undergraduate and graduate medical curricula. Other methods of evaluating clinician's (i.e., documentation in the medical record and patient feedback surveys) are not as reliable or representative of true clinician performance as USP studies [31]; a study comparing USP-facilitated audio recordings to documentation found 90% of medical record notes contained at least one error, and 3% of notes justified a higher billing level than indicated [32]. Systematically implemented, USP programs can lead to effective, accurate, and cost-efficient quality improvement, as well as the assessment of clinicians and clinical microsystem performance. In this chapter, we describe the step-by-step implementation of SPs in the clinical setting (including USPs, remote visits and ASPs), for the diverse needs of quality improvement education, research teams and health systems.

Do USPs Make Sense for Your Team? Three Requirements for Implementing USP Visits

1) A Cohort of Standardized Patients

Health professional schools that conduct OSCEs are often well-prepared to incorporate a USP program into their curriculum and assessments, as much of the infrastructure required to perform the two assessments is similar, such as SP/rater training and case development. Most academic medical centers will already have a cohort of onboarded SPs, and direct additional costs are generally limited to compensation for the actors as the encounters take place in the clinic setting. Hourly rates for standardized patients range from $15 to $30 per hour, and most USP visits last 2–4 h to allow for waiting time at the clinic and completion of post-visit surveys. Time required by administrators or faculty members to train and schedule USPs should also be factored into planning.

2) A Strong Working Relationship Between the USP Team and Clinical Leadership

USP visits can only be implemented in collaboration with members of clinical leadership who have a strong understanding of their site's clinical workflow and electronic medical record (EMR). Constant communication with scheduling

coordinators is necessary to ensure that USPs interact with the correct clinicians. Potential scheduling issues are easily overcome if the USP can book their own appointment through an online resource. Participants or leadership may be concerned that USP visits hinder productivity by taking up clinic time that could be used for real patients, but the small number of visits needed to evaluate the clinical system and clinician performance and provide actionable data is usually considered worth the investment by clinical and educational leadership.

3) A Good Attitude! There Will Be Hiccups

USP programs are not without potential difficulties. There is a risk that USPs will be detected, which can undermine the effectiveness of the program. If a medical professional realizes they are interacting with a USP, they may not behave in a natural manner. Detection rates range from 5% to 20% [33], though some studies have noted rates as high as 70% [3]. Consequently, the data collected may not reflect a clinicians' true skills. On the other hand, our work indicates most detections occur after the visit and the performance of clinicians who detect the USP is not significantly different than those who do not. Matching USPs with the intended clinician can be complicated in some health centers, particularly those that do not assign patients to specific clinicians. Nevertheless, we believe that even if a clinician identifies a USPs, it is still a valuable opportunity to receive feedback on the care one provides.

Implementing a USP Program

USP projects can be a major undertaking, and as with most educational projects, collaboration within and across specialties or disciplines can only enrich the process. While it is necessary to have strong leaders who believe in the benefits of such comprehensive assessment programs, many other individuals are needed for adequate planning, preparation, and implementation. Table 4.1 details the additional roles that USP projects typically require beyond those detailed for OSCE administration in Table 2.1. Specific tasks involved in planning to integrate USP visits in a clinical setting are detailed in Table 4.2, a modification of the worksheet for assigning OSCE responsibilities and creating timelines (Table 2.5, Appendix C).

Table 4.1 USP staffing needs

Roles	Key characteristics	# Needed
Leader	• Strong motivation to develop and implement project • Well connected to procure resources • Capable of initiating establishment of collaborative relationship with hospital/clinic leadership • Able to communicate well and create a team spirit	One or more
Planner	• Understands logistics of implementing USP assessments (case development, project location) • Can entertain multiple options for solving problems	One or more
Coordinator	• Adept at USP-related coordination (e.g., scheduling, SP recruitment, data entry) • Able to communicate well • Good at troubleshooting and problem-solving	One or more (depending on scope)
Clinical administrator	• Permitted to obtain fake medical records • Able to assess workflow to incorporate USP with no detection	Usually one
Trainer	• Understands USP roles and case requirements • Has teaching skills (e.g., provides constructive feedback) and can manage psychosocial impact of case portrayals • Able to communicate well and create a team spirit • Is sensitive to the special stresses inherent in USP work	One or more (depending on scope)
USPs	• Committed to standardization of their case portrayal (i.e., not expressing their personal creativity) • Comfortable enacting their particular medical case (i.e., not getting too involved emotionally) • Interested in taking on "educational" responsibilities • Able to tolerate the open-ended nature of USP visits (can last from 30 min to 3 h or more) • Comfortable to be among individuals who have true medical conditions and may be in emotional or physical distress (e.g., heart attack in an emergency room) • Able to change appearance if using one clinical site • Clear about USP goals and performance standards • Committed to fair performance assessments (e.g., understands personal rater style and biases) • Effective at providing post-encounter feedback	At least two per case, consider cross-trained alternates

(continued)

Table 4.1 (continued)

Roles	Key characteristics	# Needed
Data manager	• Competent at managing performance data sets • Understands USP process • Committed to accuracy	At least one
Data analyst	• Understands USP process • Has psychometric skills • Understands audience for results (e.g., learners, leadership)	At least one
Program evaluator	• Understands USP process • Is familiar with evaluation models (e.g., pre−/post-testing) • Can develop and analyze program evaluations (e.g., surveys, focus groups)	At least one

Table 4.2 Breakdown and timeline of USP project responsibilities

Initial Planning	
3–4 months before planned start of USP project	• Obtain permission from and initiate partnership with clinic leadership
	• Decide on format (e.g., number of cases, time frame) and modality (virtual or in-person)
	• Create a blueprint (identify competencies to be assessed)
	• Develop cases and gold standards/best practices of care
	• Determine who will receive report of data (learner, program leadership, clinic, etc.)
	• Identify locations (single or multiple)
	• Recruit staff for administrative tasks and scheduling
	• Identify each step of USP visit based on steps of regular patient visit (check-in procedure/sign-on procedure, insurance, medical record, etc.)
	• Decide on USP recruitment and training schedule
	• Communicate with learners to explain the nature of the project and get consent for USP visits
	• Clarify budget (e.g., USP costs, recording equipment)
Materials and visit preparations	
3 months–1 week before first USP visit	• Develop USP materials (e.g., USP instructions, rating forms)
	• Recruit USPs
	• Create medical records and unique case demographic information
	• Prepare props (e.g., fake pill bottles, inhalers, charts, insurance cards)
	• Train USPs
	• Organize practice visits (“dress rehearsals”)
	• Consider videotaping or audio-recording USP training sessions
	• Create schedule for clinic visits
	• Send demographic info (name, address, DOB) to clinic director and USP
	• Provide USP with props, audio recorder, and transportation funds (if necessary)
Visit administration	
Day of USP visit	• Send USP a reminder
	• Provide rating form for post-visit
	• Debrief with USP after the visit with the help of the rating form (consider audio- or video-taping debriefing session)
	• Plan periodic group debriefing sessions with USPs to share experiences and control for desirable and undesirable case adjustments
Post-visit tasks	
Days to weeks after USP visit	• Organize rating forms and clinic materials by case
	• Arrange for USP payment
	• Conduct a chart review of medical record
	• Survey learners for detection
	• Report evaluation data (e.g., report cards)
	• Organize materials for future reference (e.g., forms, videos)
	• Report on experience internally and externally (e.g., presentations, articles)

Cases

USP and OSCE case development follow the same basic principles (see Chap. 2, Step 4, "Develop Case Scenarios and Instructions"). Many OSCE cases can easily be adapted to be used in USP visits. To prevent detection, it is crucial to ensure that USP cases are representative of the patient population served by the clinicians being evaluated. A sample USP case and corresponding checklist, designed for an urban community clinic, are included at the end of this book in Appendices N and O. Collaboration with EMR leadership allows for clinical information to be preset in the USP's chart for construction or a patient that has already been seen at the clinic. However, USPs are usually introduced as new patients to eliminate the need to create a prior medical record in the health system's EMR.

Recruitment

The number of USPs required depends on the number of cases in the program, the number of clinicians involved, and the duration of the program. At a minimum, it is necessary to train two USPs per case. This ensures greater flexibility when scheduling visits and provides a cushion for the program if a USP suddenly drops out of the program. Recruiting several actors for each case also decreases the possibility of detection. If a USP enters the clinic frequently, they may be recognized by learners and staff. For shorter projects, telemedicine projects, or projects where you can easily predict the visit schedule, fewer are needed. Be sure to ask SPs about their availability when you interview them.

Medical schools are the best places to recruit USPs, since they work with actors who already have experience as standardized patients. The most qualified standardized patients will possess acting talent, punctuality, communication skills, and the ability to adapt to unpredictable situations.

Training

USP training sessions are similar to OSCE training exercises. Trainings can be divided into three sessions. During the first session, the USP coordinator explains the purpose and logistics of the program to USPs. USPs should then read the case instructions aloud with the USP coordinator. After the USPs fully understand their role, they practice the case, taking on the patient role, while the coordinator assumes the role of physician. The second training focuses on teaching USPs to complete the evaluation forms. The coordinator shows a presentation about the correct way to observe, categorize, and document clinicians' behavior. To practice completing the evaluation, the USPs should watch OSCE encounters and evaluate learners' skills. During the final training session, the USPs can role-play the case with an attending physician or chief resident to learn the pacing of a medical interview. The USP coordinator should also discreetly bring the USPs to the clinic before their first visit so that they can run the case with an attending in the clinic.

Clinic Location and Visits

Before any visits are planned, program leaders must get permission from clinic leadership to conduct the program. Leaders should determine how to notify learners of their participation in the project without revealing details about the USPs or specific visit timing. Several of our USP projects involving attending physicians have been opt-out and anonymous, which can encourage participation if participants are wary of being evaluated by USPs. Program leaders should also speak with members of the finance and EMR staff to learn how to prevent USP visits from being billed as real visits and appearing in clinic audits. A successful USP program results from a true collaboration between clinical and educational leadership and provides actionable feedback relevant to current priorities and quality initiatives to the clinicians and the health system.

The USP coordinator should visit the clinic during a busy day to observe its layout. They should note the location of the registration desk, exam rooms, and other relevant areas (finance desk, pharmacy, etc.) The USP coordinator must observe where patients must go to check in, pay, encounter doctors, and get prescriptions and referrals; clinical leadership can also help clarify this information.

After visiting the clinic, the program team should identify unique characteristics in workflow or environment that might influence USP visits. In some clinics, it is possible to schedule appointments with a specific doctor; in others, particularly for new patients, patients are assigned doctors on a first come, first serve basis. The USP coordinator needs to work with a clinical administrator to develop a system that will ensure USPs are sent to the correct physicians. A clinical administrator should also be responsible for entering USPs' demographic information into the clinic's computer system. The USP coordinator can develop a process and deadline for sending the demographic information for each visit.

Scheduling USPs requires collaboration and planning on the part of the clinical administrator, the USP coordinator, and the USPs. If there is no online scheduling system, the USP coordinator chooses dates and times for USP visits after confirming USP availability and sends them to the clinic administrator for approval in advance of the planned visits. The clinic administrator approves the requests if the appointments are available and can suggest edits to the USP coordinator's selections if there are scheduling conflicts.

After the schedule is finalized, the USP coordinator confirms the visits with the USPs. If the clinic has the capacity for online scheduling, the USP can also sign up for the visits directly, allowing them to evaluate the visit scheduling process. This protocol will vary depending on the clinical setting and EMR system.

Practices often call or text USPs prior to and following the visit to remind them of appointments and schedule follow-up care. It may be necessary to use a generated phone number in the chart to conceal the USP's real identity and capture all pre- and post-visit contact.

If permission has been obtained from clinical leadership, audio can be recorded in order to validate the checklist data during the clinical encounters. USPs can use their iPhone or a small recorder in their pocket or purse.

The USP team must develop a plan to excuse the USP from labs, blood work, or additional time in the hospital immediately following the visit. It is usually effective for the USP to have a brief excuse to avoid getting labs ordered by a physician; they might "have no time because they need to go back to work" or they "just don't want to today." USPs should hold on to any paperwork they are given (prescriptions, referrals, after visit summaries, etc.) and return it to the USP coordinator after their visit.

Post-visit

After the visit is complete, the USP will complete the evaluation checklist of the visits. This can be done on a phone, iPad, or computer either on site (if the visit was near the USP coordinator's office/desk) or off-site (if the USP visit was not near the coordinator). It is also ideal for the coordinator to debrief the visit with the USP, either in-person or on the phone. During this discussion, logistical issues/challenges can be noted, and the USP can share additional qualitative data. Topics raised during the debriefing include the atmosphere of the clinic, the conduct of the clinician and/or medical assistants, and the degree of difficulty in navigating the clinic. Debriefing sessions should also explore facilitators and barriers to patient care. The USP coordinator should track unexpected occurrences during the visit or changes in the hospital system that might impact future USP visits.

Alternative Model: Remote USP Visits

While this chapter outlines the implementation of USPs for in-person visits with a clinician, it is relatively simple to modify the model to evaluate virtual clinical encounters. Given the COVID-19-induced rapid increase in remote (video or audio) care in the spring of 2020, the ability to assess virtual visits for quality control and quality improvement is crucial for modern health systems. USPs present a useful strategy for this work. Key modifications to the process are noted below.

Cases

Because virtual visits require distinct clinical skills related to the use of technology and the ability to observe or explore the patient's home environment, USP cases should be designed to adequately evaluate these skills, for example, an older adult case with obvious tripping hazards in the background or a patient who cannot remember the name of her medication but, if asked, can go to the bathroom and read or show the prescription bottle. Sample telemedicine checklist items can be found in Chap. 9, Table 9.1.

Training

While the initial two trainings can be conducted in person or remotely, the final training with a clinician should be conducted virtually to ensure the USP case is conveyed appropriately over video/audio and the video background is appropriate. The USP coordinator should schedule a test visit with a clinician so the USP can try out the process of signing in and out of an actual virtual visit.

Location of Visits

Remote visits can be conducted by the USP in their own home. When training the USP, be sure to identify the most appropriate, realistic location in their home for them to conduct the remote visit. In addition to the standard demographic information for each visit (name, DOB, address), the USP will also need a fake EMR log-in which the clinical administrator should be able to provide.

Post-visit Debrief

At the end of the visit, the USP should debrief with the USP coordinator on the phone, complete the checklist online, and then alert the USP coordinator that that the visit is complete.

Chart Review for Remote and in-person USP visits provides additional information

Review of the clinical note that the clinician writes during the USP encounter can provide additional information on their skills. The USP program team should develop a chart review checklist that includes items to assess how the clinical record captures the physical exam, patient medical his-

tory, treatment plan, and follow-up care including any prescriptions, labs, or referrals. There may be significant overlap between USP checklist items and chart review items. The chart review does not need to be completed by a clinician; with proper training, it can be completed by any member of the USP team. Chart review data will allow you to compare what the clinician does as reported by the USP and what the clinician documents in the medical record, as well as how well the system responded to their charting.

Depending on which EMR system is used, the export or login process to access the chart will look different. It is likely easiest to have a clinician export PDFs of the charts or to have the USP coordinator secure read-only access to the EMR.

Budgeting for USPs

The USP coordinator can keep track of the program costs in an Excel spreadsheet. All training and visit costs for each USP should be documented and updated frequently to ensure the program stays within the budget. The spreadsheet should include the name and contact information of each USP, list every date each USP worked, and include the amount USPs were paid for each visit or training session. While the USP coordinator should always keep track of the USP hours, if the USPs are recruited through a medical school, it may be necessary to pay them through the medical school directly.

Alternative Model: Announced Standardized Patient (ASP) Visits

USPs provide a valuable mechanism by which to understand a patient's clinical experience; however logistics, time available, and interest from leadership can present significant barriers. Encounters using announced standardized patients (ASPs), where the clinic and clinician are informed that the patient is an actor, can be simpler to implement and provide much of the same information as USPs. For instance, ASPs allow the clinician to practice core skills in their "real environment" using the same EMR they use with real patients. ASP visits should be scheduled through the same process as a USP visit, but the clinician should be made aware that they will receive a visit at a particular time. The clinician should be instructed to treat the ASP visit as they would a regular visit and should not be informed of the details of the case. A notable benefit of using an ASP over a USP is that the ASP can provide immediate feedback to the clinician at the end of the visit before completing the standardized checklist, whereas the USP must remain in character until the visit ends. Table 4.3 outlines key differences between USP and ASP visits.

Practical Use of USP Data: Learner, Clinical Microsystem, and Programmatic Evaluation

USP visits provide a wealth of information for learners, clinical leaders, and education leaders that can facilitate the growth of a learning health system (Fig. 4.1). The SP collects data from the moment they walk into the office till they leave. From this data, the USP team can create and disseminate (1) individual reports for learners on their overall performance across the USP cases they received (Fig. 4.2) and/or (2) summary reports for educational or clinic leadership on the overall performance of their residents/learners, including information on patient safety, patient centeredness, screening assessments, and team skills (Fig. 4.3).

For further analysis of USP data, USP scores can also be compared to OSCE performance to determine how individu-

Table 4.3 USP and ASP similarities and differences

Type of visit	Unannounced standardized patient	Announced standardized patient
Cases	The same cases can be used for ASP and USP visits	
Recruitment	The same SPs can be used for ASP and USP visits	
Training	USP and ASP training for the visit is the same	ASPs should receive additional training on how to debrief the visit with the clinician at the end of the visit
Clinic location of visits	USP and ASP visits should both occur in the clinician's regular clinical space	
Scheduling visits	Visits should be scheduled with a clinic administrator who will not disclose visit to clinical team and will ensure the visit looks normal in the clinician's schedule	Visits can be scheduled in clinician's regular schedule or during protected, nonclinical time; they can be scheduled with the clinician or through the regular system
Post-visit	USP cannot provide immediate feedback so as not to "unmask" the case for future visits to other clinicians in the system	Opportunity to provide clinician with immediate feedback about their performance
Evaluation of clinician	Clinician is unaware that they are being evaluated	Clinician is aware that they are being evaluated
Evaluation of microsystem	Clinical system (front desk, medical assistant) is unaware that they are being evaluated	Clinical system (front desk, medical assistant) is aware that they are being evaluated

Provider ID:

Domain Summary Scores by Case

Communication Skills

Your overall communication score was calculated as the percentage of well done across all communication skill items, as rated by Standardized Patients. This includes all items in the 3 domains below (e.g. Information Gathering, Relationship Development and Patient Education). Performance on individual items that make up each core domain are located in the second half of the report.

Information Gathering

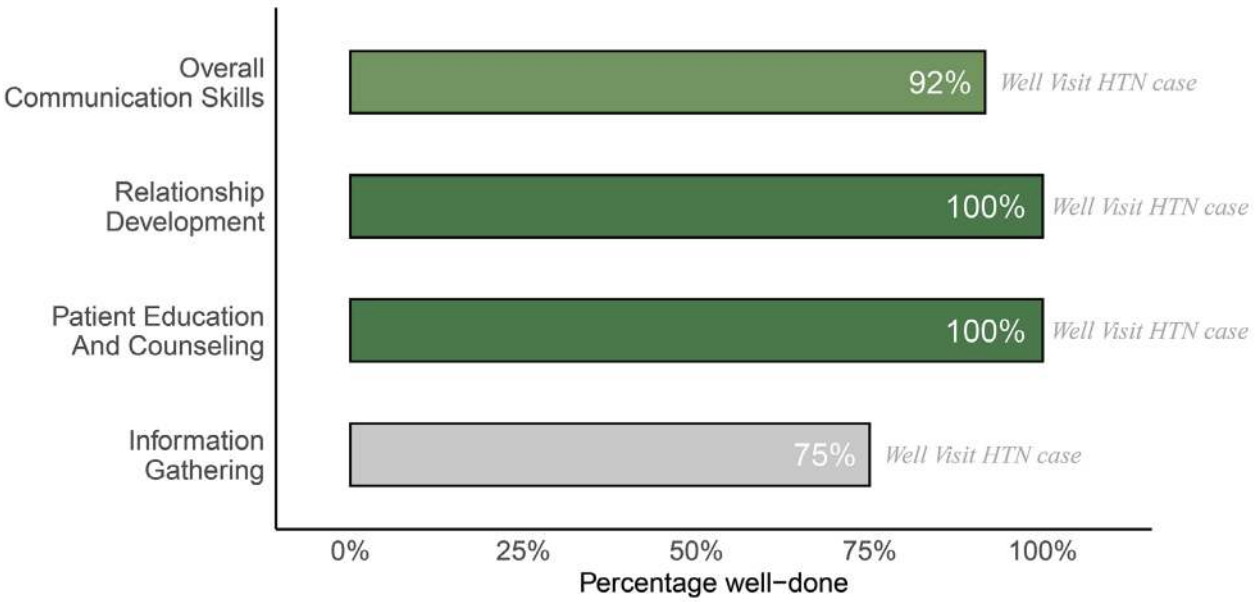
Standardized Patients rated how well you were able to manage the narrative flow during the encounter by allowing the them to talk without interruptions, asking them appropriate questions, and clarifying/repeating information throughout the encounter to ensure understanding from both sides.

Relationship Development

Standardized Patients rated how well you were able to communicate your concern/intention to help them including your non-verbal communication, and the words/jargon you used. They also considered the way you made them feel by how well you acknowledged their emotions and how accepting/nonjudgmental you were.

Education And Counseling

Standardized Patients rated how well you asked them questions about their understanding, how clear your explanations were, and how much you collaborated with them to identify next steps in their care plan.



Overall Communication Comments

Comments about the provider's communication, education, and pacing skills

Well Visit Hypertension Case

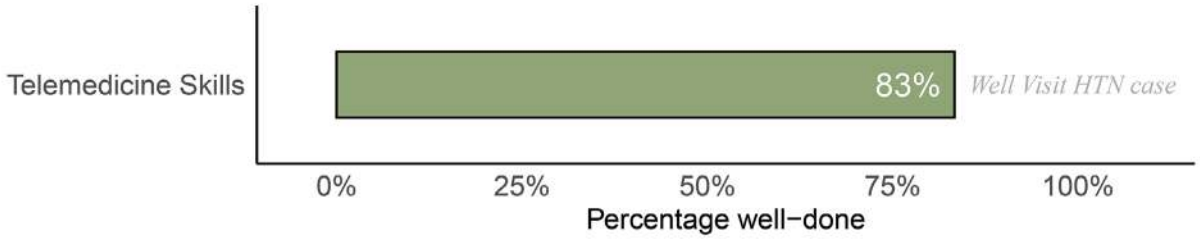
"I appreciated that the doctor helped to create a plan for my continued healthcare."

"I wish we could have discussed my current situation with my family."

"It was very helpful that the doctor mentioned different ways/medications to help stop smoking."

Fig. 4.2 Sample individual clinician USP report

Telemedicine Skills



Telemedicine Comments

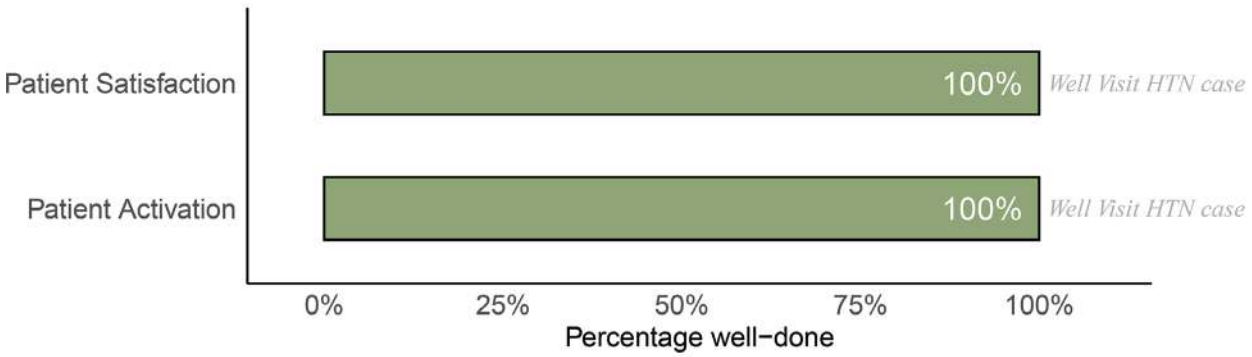
Comments about the provider's telemedicine skills

Well Visit Hypertension Case

"The doctor was great with switching to using the phone instead, once we started having issues with our video connection."

Patient Experience Skills

This includes general satisfaction, activation (the sense that they could manage their health condition) and management and treatment planning.



Patient Experience Comments

Comments about the provider's patient activation, patient satisfaction, and management and treatment skills

Well Visit Hypertension Case

"It was very helpful that the doctor mentioned different ways/medications to help stop smoking."

Global Recommendation

Standardized Patients consider their overall experience and rate whether they would recommend you to family and friends and your overall professionalism. Both items are rated on 4-point scales. Recommend: Recommend Highly, Recommend, Recommend with Reservations and Not Recommend. Professionalism: Completely professional, Mostly professional, Somewhat professional, Not at all professional.

Communication

The patient from the Well Visit Hypertension Case would **Recommend you to family and friends**.

Professionalism

The patient from the Well Visit Hypertension Case felt that you were **Completely professional**.

Fig. 4.2 (continued)

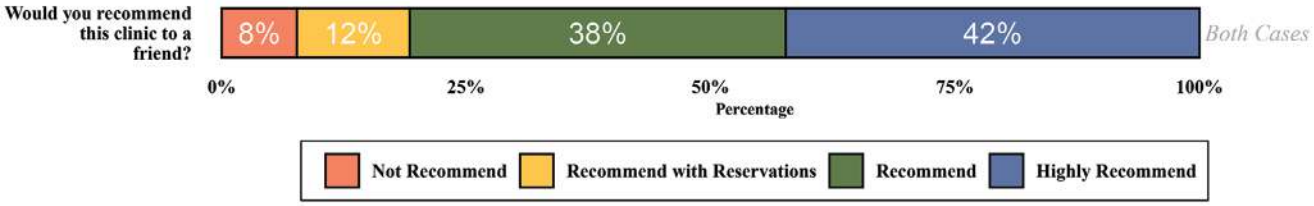
Aggregate Patient Perspective on Telemedicine Care: Targeted Clinic-Level Feedback from USP Visits

We hope you find this summary report insightful and find the data it provides on your team's clinical skills, practice, and the clinical microsystem to be valuable. Substantive, narrative commentary from USPs highlight areas of strengths and challenges. Some of these key takeaways are below.

Clinic Functioning

Overall Clinic Recommendation

Number of Visits: 26



Comments on Clinic Recommendation by Rating

Not Recommend

"After an hour waiting in the video pt waiting room, pt receives a call from the resident not clearly identifying herself and sounding frazzled. When pt mentioned that they had been waiting for an hour on the video call (she attempted to do the visit via phone) no apology was forthcoming but instead resident said that they were not the only pt they had this morning and dismissed the wait while attempting to log on to the video. In addition to Resident being dismissive, curt, abrupt and seemingly judgmental, Sound quality on the video was not ideal which added to pt's unwillingness to engage further in dialogue."

"Long Waiting time with no communication from the provider x 2."

Recommend with Reservations

"Am recommend with reservations based on the clinic's request for me to buy a BP to bring with me to have someone at the hospital check my blood pressure. This was truly the most absurd thing I've been told working this program to date."

"Provider was responsive, on time, expressed a willingness and readiness to help. Pt could benefit from being partnered with a little bit more."

"As a patient I felt a bit confused as to how to follow up with their if I needed to speak with them."

Recommend

"Doctor was warm, approachable and very professional dur-ing the technical difficulties."

"The virtual visit was very friendly and I felt like I was prescribed what I needed to help me feel better."

"The physician was very friendly and help me feel that my asthma will get better."

"The provider was very respectful and had good empathy towards my Asthma condition."

Highly Recommend

"Helpful. Intent to help really came across."

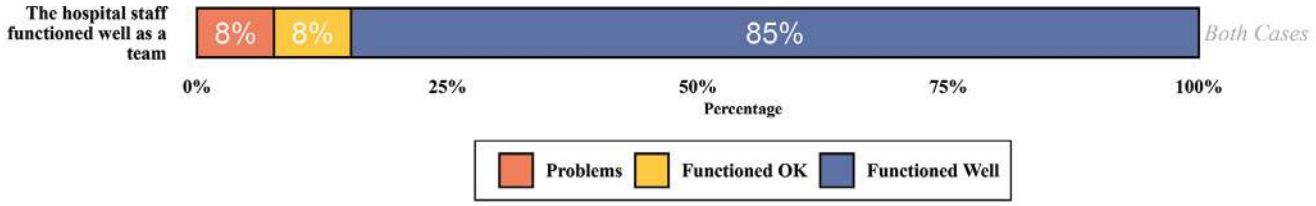
"It was fairly easy to navigate the system for this virtual visit. I think that's VERY IMPORTANT especially for this patient demo-graphic."

"The appointment was on time. The provider was clear and friendly. I also felt confident that I will get better."

"The doctor was very professional and had great connection and empathy with the patient."
"This provider was great!"

"Great empathy, they took their time with me and answered all of my questions."

Fig. 4.3 Sample clinic-level USP report



Comments on Clinic Team Functioning by Rating

Problems

"A nurse called and scheduled 2 follow up appointments. I was called moments later and was told the hospital did not have any more BP machines, and that I would have to buy one. The nurse scheduled and appointment for me to come into the hospital to have my BP measured. I was told I would have to buy a BP machine, and bring the machine to the hospital in order to have it checked. Which made no sense. I asked the nurse why I needed to buy a BP to bring to the hospital, assuming there are literally hundreds of BP machines in the hospital to check a patients BP. They thought it was odd too, and could not come up with a reason for such an odd request."

"The team seemed busy and somewhat dismissive and insensitive to the pt."

Functioned OK

"The call was 45 mins late but the handoff and promised call backs were prompt."

"Someone knocked on their door and interrupted while we were having our phone visit."

"Asked only for my name but did not confirm with birthdate - both RN and Provider."

Functioned Well

"both the doctor and nurse contacted me in a timely fashion. The nurse called me a few mins after I was done speaking with the doctor."

"The doctor mentioned that someone would call me to follow up from our conversation. Not long after our encounter, I did actually receive a call

"Only interacted with the Provider."
"

The nurse called me shortly after my telehealth call with the Dr."

"I received a callback right away from the nurse to schedule my follow up appt."

"Did not meet or have contact with anyone else on the team ."

"Simple, easy but the long wait for a follow up phone appt gave pause."

Follow Up Care

Did you receive information about ...

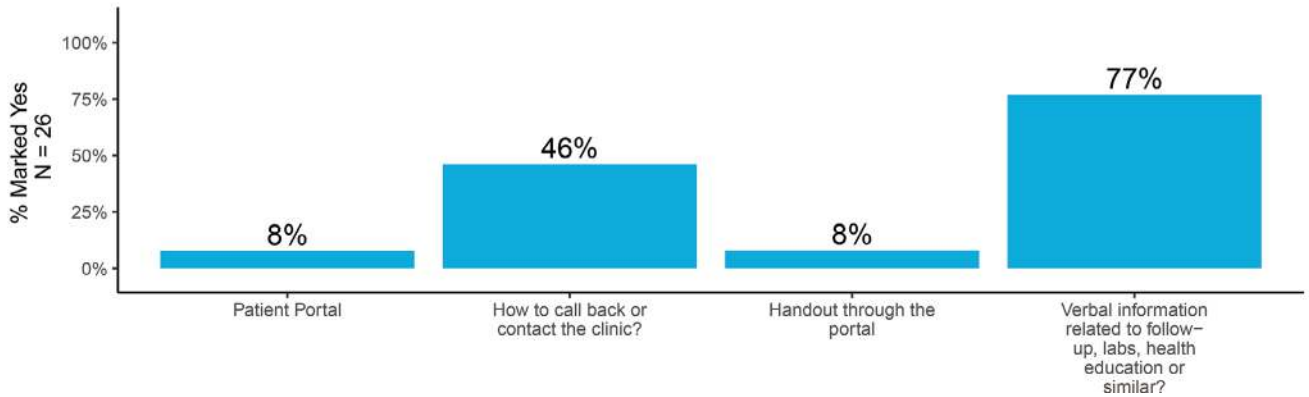


Fig. 4.3 (continued)

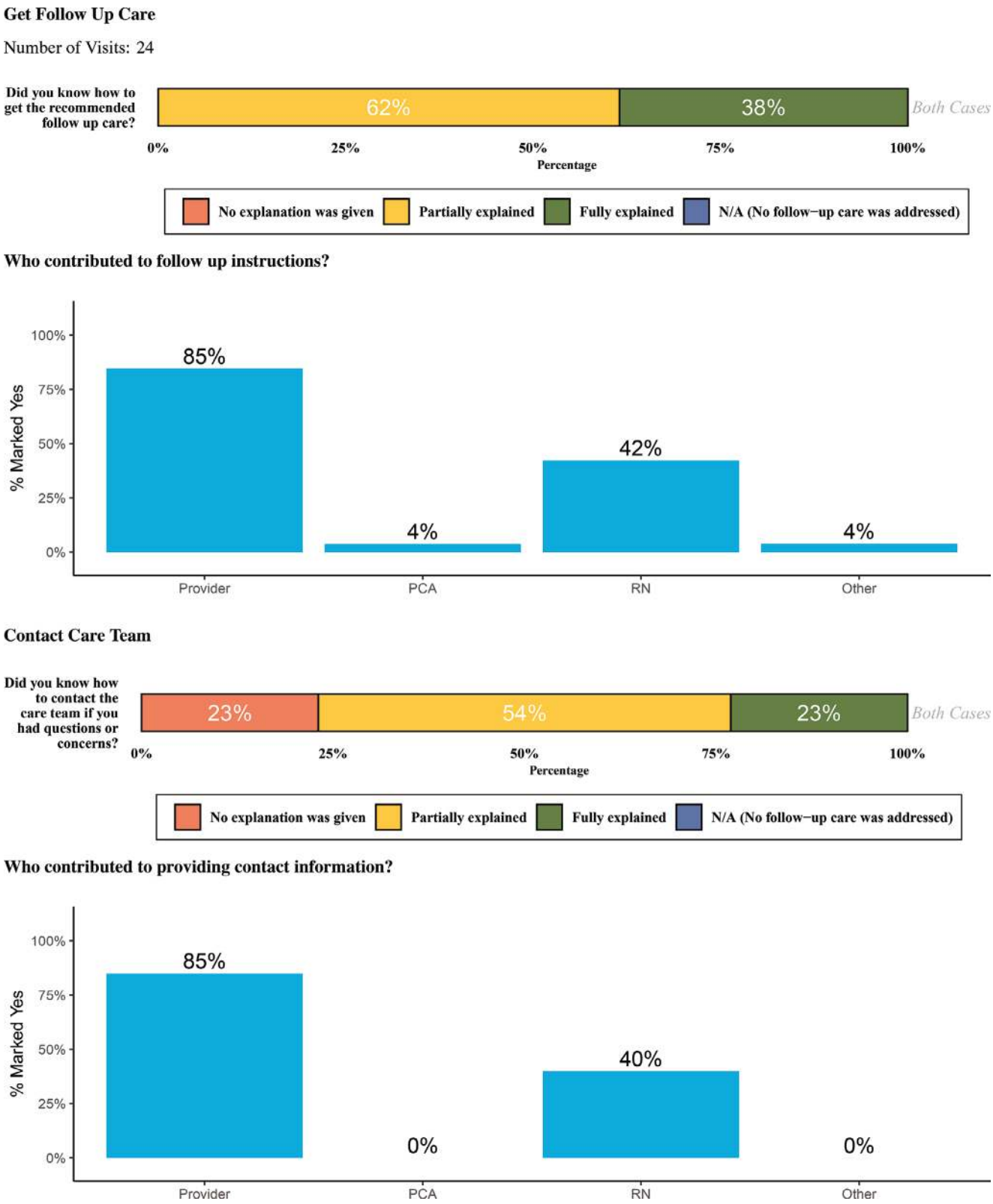


Fig. 4.3 (continued)

	% Well Done (N=14)	
Aggregate Patient Experience of Provider Care	**Item Score**	**Summary Score (Mean)**
Managed the narrative flow of your story	79%	77%
Clarified information throughout by repeating to make sure they understood you	71%	
Allowed you to talk without interrupting	93%	
Relationship Development		
Communicated concern or intention to help	86%	86%
Non-verbal behavior enriched communication (e.g., eye contact, posture)	80%	
Acknowledged emotions/feelings appropriately	79%	
Was accepting/non-judgmental	93%	
Used words you understood and/or explained jargon	93%	
Education and Counseling		
Asked questions to see what you understood (checked your understanding)	14%	48%
Provided clear explanations/information	79%	
Collaborated with you in identifying possible next steps/plan	50%	

SDOH Housing

Out of the **14** completed visits, **2** reported that the clinician asked about/explored their housing situation. Of those 2, both "acknowledged the patient's concern" but did not fully explore.

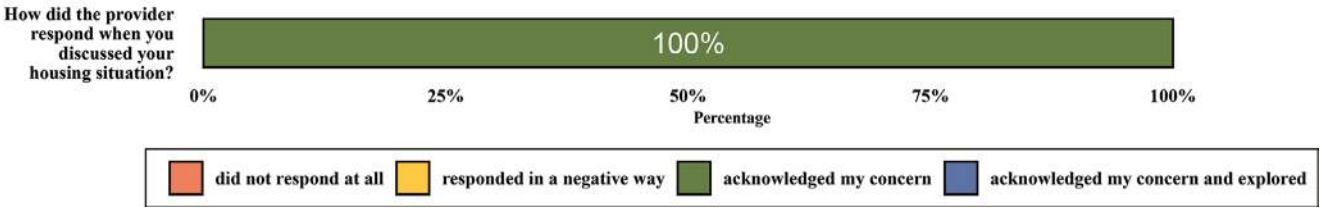


Fig. 4.3 (continued)

Labs	Number of Orders (N= 14)
Labs Ordered	
HIV test	9
Lipids lab	1
A1c (for diabetes) lab	8
Other tests/labs	8
COVID Vaccine	10
Flu Vaccine	2
TDAP	2
A1C	1
Chlamydia	6
Syphilis	6
Basic Metabolic Panel	3
Urinalysis	1
Complete Blood Count (CBC)	11
Pulmonary fit test (PFT)	2
Hep B test	6

Hypertension

	% Marked Yes (N=14)
Hypertension Items	
Were you schedule for a follow up visit regarding hypertension?	100%
Did the physician ask you if you have a home blood pressure monitor?	79%
Did they tell you that you can get a free blood pressure monitor from the clinic/ pharmacy?	50%

If scheduled for a follow up visit regarding hypertension, which do you think you were scheduled for?

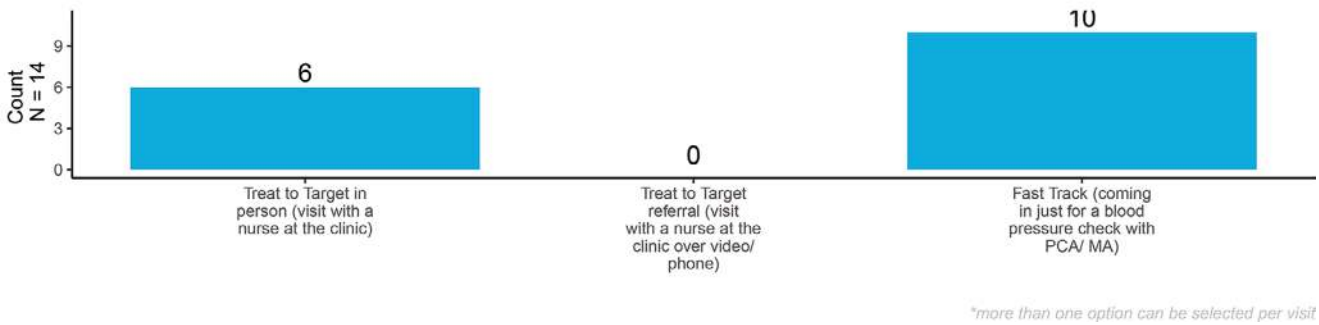


Fig. 4.3 (continued)

als perform in a testing situation (OSCE) vs. the "real world" (USP) in the clinical environment. The sample evaluation in Fig. 4.4 shows resident communication skills by individual and mean cohort as measured in an eight-station OSCE and across multiple USP visits (% checklist items "well done"). The particular resident in the sample report ("Dr. K") actually shows a trend of performing better in USP visits as compared with OSCE encounters.

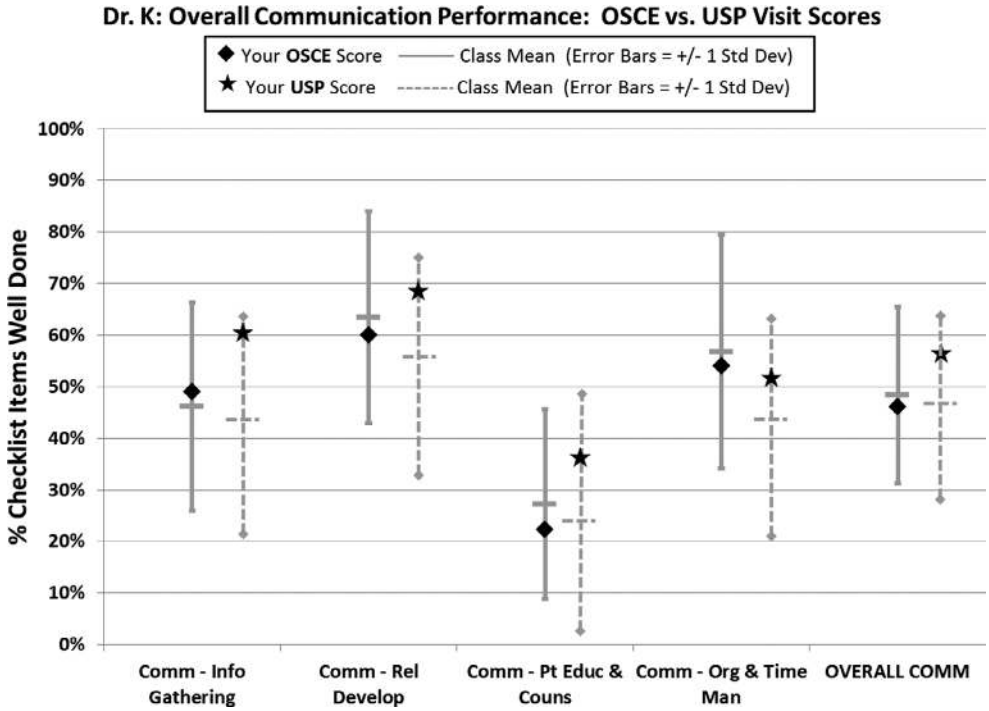


Fig. 4.4 Sample learner feedback report: OSCE vs. USP communication performance. Communication sub-competency scores reported: information gathering, relationship development, patient education and counseling, and organization and time management

Our Experience with USP Outcomes and the Learning Health System

As defined by the Agency for Health Care Research and Quality (AHRQ), a learning health system is "a health system in which internal data and experience are systematically integrated with external evidence, and that knowledge is put into practice." This enables healthcare systems to provide higher quality, safer and more efficient care (Fig. 4.1). USPs are a mechanism of establishing a learning health system as USPs quickly and systematically collect data and information about their experiences, which can be fed back to clinical leadership.

In our program, USPs evaluate clinical microsystem functioning as well as learner performance (residents and attending clinicians) at urban public ambulatory care clinics, private primary care clinics, and a private college health center, among others. During each visit, USPs evaluate learners' case-specific performance and the broader clinical system. These items include whether the medical assistants greet the USP within a reasonable time frame; introduce themselves; wear a visible name tag; wash hands before touching the USP; measure the USP's height, weight, and blood pressure; and screen for depression. All USPs use the same behaviorally anchored checklist to assess clinician performance and the general experience of the clinic: ease of navigating the system, team functioning, and overall staff professionalism. Then, we create reports to feed this information back to the clinical or educational leadership, who in turn can create systems or curricular change.

Many of our projects focus on system-wide assessment and quality improvement. From 2011 through 2020, our team facilitated 987 USP visits with residents associated with our institution. When the COVID-19 pandemic restrictions shifted care to telemedicine, the need for comprehensive telemedicine evaluations and trainings became clear. Our team worked quickly to implement 2 telehealth USP programs with 104 remote USP visits conducted over video or phone. Other projects have explored system competency for addressing specific aspects of the visit, such as screening for social determinants of health [24] or depression [25], and preparedness for emerging health challenges, such as the COVID-19 pandemic [22] and teaching telemedicine skills [34]. Table 4.4 lists some of our USP programs' actionable recommendations.

Table 4.4 Sample USP programs and findings

USP program and reporting structure	Issue identified	Actionable recommendation
Program: USP visits ($n = 72$) evaluate telemedicine care at three urban ambulatory care clinics **Reports/feedback**: Aggregate, de-identified report to clinical leadership team; individual report to clinician	(1) System issues related to the remote platform (2) Clinics struggling to integrate the clinical care team into the telehealth visit (e.g., 36% of USPs received a pre-visit call) (3) Clinicians were burdened with most screening and follow-up tasks (e.g., only 5% were screened for depression)	• Establish standardized call procedures to collect screening and insurance information in addition to reminder notifications • Determine differences between pre-visit, portal, and in-visit screenings • Define the role of the care team in conducting screenings
Program: USP visits ($n = 4$) assess a clinic's COVID-19 preparedness and ability to rapidly identify and isolate potentially infected individuals at onset of pandemic **Reports/feedback**: Aggregate, de-identified report to clinical leadership team, USP discloses identity after clinical team isolates them and then provides immediate feedback	(1) In half of the visits, frontline staff either failed to recognize or solicit triggers that would have merited immediate isolation of the patient (2) In half of the visits, frontline staff told the USP to go to another floor or return to the waiting room until further notice *without providing a mask or calling a nurse or other clinician to perform a clinical assessment*	• Clarify to front desk staff protocol for potentially infected individual • Staff feedback informed necessary protocol changes for handling patients under investigation in future encounters
Program: USP visits ($n = 85$) assess clinic screening and internal medicine (IM) residents' practices in identifying, documenting, and treating depression **Reports/feedback**: Aggregate, de-identified report to residency leadership team; individual report to resident	(1) Only 85 residents (66%) appropriately referred the USP (2) 79 residents (61%) provided guideline-appropriate treatment (3) 59 residents (46%) prescribed an SSRI (4) 49 (38%) scheduled a 2-week follow-up (5) 40 (31%) did not add depression to the problem list	• Clarify referral protocol for patients with positive PHQ-9 • Refresh preceptors on appropriate follow-up visit timeline for patient with concern for depression

Conclusion

USPs offer great promise for overcoming methodological issues often associated with OSCEs, ASPs, and other traditional assessment methods. Trained to be reliable and unbiased raters of clinical performance and clinic functioning, USPs provide feedback that reflects the actual practice of healthcare systems. USPs offer unique insight into the perspective of the patient and avoid the biases of patient surveys and the "Hawthorne effect," or the altered behavior of clinicians in OSCEs due to their awareness of being observed. Though possible detection of USPs and concerns about the ethics of sending "fake" patients to clinics may limit the effectiveness of a USP program, the structured feedback provided by USPs can be exceptionally valuable. We highly recommend implementing a USP program to measure and improve health systems.

References

1. Ozuah PO, Reznik M. Residents' asthma communication skills in announced versus unannounced standardized patient exercises. Ambul Pediatr. 2007;7(6):445–8. https://doi.org/10.1016/j.ambp.2007.06.005
2. Glassman PA, Luck J, O'Gara EM, Peabody JW. Using standardized patients to measure quality: evidence from the literature and a prospective study. Jt Comm J Qual Improv. 2000;26(11):644–53. https://doi.org/10.1016/S1070-3241(00)26055-0
3. Rethans JJ, Gorter S, Bokken L, Morrison L. Unannounced standardised patients in real practice: a systematic literature review. Med Educ. 2007;41(6):537–49. https://doi.org/10.1111/j.1365-2929.2006.02689.x
4. Carney PA, Ward DH. Using unannounced standardized patients to assess the HIV preventive practices of family nurse practitioners and family physicians. Nurse Pract. 1998;23(2):56–8.
5. Derkx HP, Rethans JE, Maiburg BH, Winkens RA, Muijtjens AM, Van Rooij HG, Knottnerus JA. Quality of communication during telephone triage at Dutch out-of-hours centres. Patient Educ Couns. 2009;74(2):174–8. https://doi.org/10.1016/j.pec.2008.08.002
6. Lee L, Burnett AM, D'Esposito F, Fricke T, Nguyen LT, Vuong DA, Nguyen HTT, Yu M, Nguyen NVM, Huynh LP, Ho SM. Indicators for assessing the quality of refractive error care. Optom Vis Sci. 2021;98(1):24–31. https://doi.org/10.1097/OPX.0000000000001629
7. Shah R, Edgar D, Evans BJ. Measuring clinical practice. Ophthalmic Physiol Opt. 2007;27(2):113–25. https://doi.org/10.1111/j.1475-1313.2006.00481.x
8. Culver JO, Bowen DJ, Reynolds SE, Pinsky LE, Press N, Burke W. Breast cancer risk communication: assessment of primary care physicians by standardized patients. Genet Med. 2009;11(10):735–41. https://doi.org/10.1097/GIM.0b013e3181b2e5eb
9. Schwartz A, Peskin S, Spiro A, Weiner SJ. Direct observation of depression screening: identifying diagnostic error and improving accuracy through unannounced standardized patients.

Diagnosis (Berl). 2020;7(3):251–6. https://doi.org/10.1515/dx-2019-0110
10. Chung AS, Bogoch S, Mody S, et al. Videotaped unannounced standardized patient encounters to evaluate interpersonal and communication skills in emergency medicine residents. AEM Educ Train. 2019;4(4):419–22. https://doi.org/10.1002/aet2.10401
11. Zabar S, Ark T, Gillespie C, Hsieh A, Kalet A, Kachur E, Manko J, Regan L. Can unannounced standardized patients assess professionalism and communication skills in the emergency department? Acad Emerg Med. 2009;16(9):915–8. https://doi.org/10.1111/j.1553-2712.2009.00510.x
12. Mohsin A, Williams MD. No-notice mystery patient drills to assess emergency preparedness for infectious diseases at community health centers in New York City, 2015–2016. J Community Health. 2019;44(2):387–94. https://doi.org/10.1007/s10900-018-00595-5
13. Ozuah PO, Reznik M. Using unannounced standardized patients to assess residents' competency in asthma severity classification. Ambul Pediatr. 2008;8(2):139–42. https://doi.org/10.1016/j.ambp.2007.11.004
14. Casebeer LL, Klapow JC, Centor RM, et al. An intervention to increase physicians' use of adherence-enhancing strategies in managing hypercholesterolemic patients. Acad Med. 1999;74(12):1334–9. https://doi.org/10.1097/00001888-199912000-00018
15. Shirazi M, Lonka K, Parikh SV, Ristner G, Alaeddini F, Sadeghi M, Wahlstrom R. A tailored educational intervention improves doctor's performance in managing depression: a randomized controlled trial. J Eval Clin Pract. 2013;19(1):16–24. https://doi.org/10.1111/j.1365-2753.2011.01761.x
16. Wilhite JA, Fisher H, Altshuler L, Cannell E, Hardowar K, Hanley K, Gillespie C, Zabar S. Gasping for air: measuring patient education and activation skillsets in two clinical assessment contexts. BMJ Simul Technol Enhanc Learn. 2020a;7(5):428–30. https://doi.org/10.1136/bmjstel-2020-000759
17. Wilhite JA, Hardowar K, Fisher H, Porter B, Wallach AB, Altshuler L, Hanley K, Zabar SR, Gillespie CC. Clinical problem solving and social determinants of health: a descriptive study using unannounced standardized patients to directly observe how resident physicians respond to social determinants of health. Diagnosis (Berl). 2020b;7(3):313–24. https://doi.org/10.1515/dx-2020-0002. Erratum in: Diagnosis (Berl). 2020 Oct 27;8(1):129. PMID: 32735551.
18. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. JAMA. 2000;283(13):1715–22. https://doi.org/10.1001/jama.283.13.1715
19. Kunow C, Langer B. Using the simulated patient methodology to assess the quality of counselling in German community pharmacies: a systematic review from 2005 to 2018. Int J Pharm Sci. 2021;13(1):10–9. https://doi.org/10.22159/ijpps.2021v13i1.40034
20. Rashid ZA, Al-Tabakha MM, Alomar MJ. Proper counseling and dispensing of Isotretinoin capsule products by community pharmacists in UAE: a simulated patient study. Clin Cosmet Investig Dermatol. 2020;13:405–14. https://doi.org/10.2147/CCID.S256302
21. Stange KC, Nutting PA, Miller WL, Jaén CR, Crabtree BF, Flocke SA, Gill JM. Defining and measuring the patient-centered medical home. J Gen Intern Med. 2010;25(6):601–12. https://doi.org/10.1007/s11606-010-1291-3
22. Fisher H, Re C, Wilhite JA, et al. A novel method of assessing clinical preparedness for COVID-19 and other disasters. Int J Qual Health Care. 2021;33(1):mzaa116. https://doi.org/10.1093/intqhc/mzaa116
23. Epstein RM, Franks P, Shields CG, et al. Patient-centered communication and diagnostic testing. Ann Fam Med. 2005;3(5):415–21. https://doi.org/10.1370/afm.348
24. Gillespie C, Wilhite JA, Hanley K, Hardowar K, Altshuler L, Fisher H, Porter B, Wallach A, Zabar S. Addressing social determinants of health in primary care: a quasi-experimental study using unannounced standardised patients to evaluate the impact of audit/feedback on physicians' rates of identifying and responding to social needs. BMJ Qual Saf. 2023;32(11):632–43. https://doi.org/10.1136/bmjqs-2021-013904
25. Zabar S, Hanley K, Watsula-Morley A, Goldstein J, Altshuler L, Dumorne H, Wallach A, Porter B, Kalet A, Gillespie C. Using unannounced standardized patients to explore variation in care for patients with depression. J Grad Med Educ. 2018;10(3):285–91. https://doi.org/10.4300/JGME-D-17-00736.1
26. Hanley K, Zabar S, Altshuler A, Lee H, Ross J, Rivera N, Marvilli C, Gillespie C. Opioid vs nonopioid prescribers: variations in Care for a Standardized Acute Back Pain Case. Subst Abus. 2017;38(3):324–9. https://doi.org/10.1080/08897077.2017.1319894
27. Kelley AT, Smid MC, Baylis JD, Charron E, Binns-Calvey AE, Archer S, Weiner SJ, Begaye LJ, Cochran G. Development of an unannounced standardized patient protocol to evaluate opioid use disorder treatment in pregnancy for American Indian and rural communities. Addict Sci Clin Pract. 2021;16(40) https://doi.org/10.1186/s13722-021-00246-6
28. Schwartz A, Peskin S, Spiro A, Weiner SJ. Impact of unannounced standardized patient audit and feedback on care, documentation, and costs: an experiment and claims analysis. J Gen Intern Med. 2021;36(1):27–34. https://doi.org/10.1007/s11606-020-05965-1
29. Wilhite JA, Velcani F, Watsula-Morley A, Hanley K, Altshuler L, Kalet A, Zabar S, Gillespie CC. Igniting activation: using unannounced standardized patients to measure patient activation in smoking cessation. Addict Behav Rep. 2019;9:100179. https://doi.org/10.1016/j.abrep.2019.100179
30. Institute of Medicine (US). Roundtable on value & science-driven health care. In: Yong PL, Olsen L, McGinnis JM, editors. Value in health care: accounting for cost, quality, safety, outcomes, and innovation. Washington, DC: National Academies Press (US); 2010.
31. Weiner SJ, Schwartz A. Directly observed care: can unannounced standardized patients address a gap in performance measurement? J Gen Intern Med. 2014;29(8):1183–7. https://doi.org/10.1007/s11606-014-2860-7
32. Dunsch F, Evans DK, Macis M, Wang Q. Bias in patient satisfaction surveys: a threat to measuring healthcare quality. BMJ Glob Health. 2018;3(2):e000694. https://doi.org/10.1136/bmjgh-2017-000694
33. Weiner SJ, Wang S, Kelly B, Sharma G, Schwartz A. How accurate is the medical record? A comparison of the physician's note with a concealed audio recording in unannounced standardized patient encounters. J Am Med Inform Assoc. 2020;27(5):770–5. https://doi.org/10.1093/jamia/ocaa027
34. Franz CE, Epstein RM, Miller KN, Brown A, Song J, Feldman M, Franks P, Kelly-Reif S, Kravitz RL. Caught in the act? Prevalence, predictors, and consequences of physician detection of unannounced standardized patients. Health Serv Res. 2006;41(6):2290–302. https://doi.org/10.1111/j.1475-6773.2006.00560.x
35. Phillips Z, Wong L, Crotty K, Horlick M, Johnston R, Altshuler L, Zabar S, Jay M, Dembitzer A. Implementing an experiential telehealth training and needs assessment for residents and Faculty at a Veterans Affairs Primary Care Clinic. J Grad Med Educ. 2023;15(4):456–62. https://doi.org/10.4300/JGME-D-22-00868.1

5 Post-encounter Learning and Assessment of Clinical Reasoning: Patient Notes and Reflections

Verity Schaye, Ruth Crowe, Christine Beltran, and Colleen Gillespie

Although some skills of clinical reasoning can be assessed with the OSCE patient interview and physical examination, a post-encounter patient note and reflection affords the opportunity for the learner to demonstrate the ability to organize information gathered, synthesize and interpret data, and document a prioritized differential diagnosis and management plan.

Background

Clinical reasoning is a complex construct with multiple components including information gathering, hypothesis generation, forming a problem representation (a summary of the patient presentation using abstract terms, unified medical concepts, and semantic qualifiers), formulating and prioritizing a differential diagnosis with diagnostic justification, and finally, developing a management and treatment plan [1, 2, 3]. Assessment of clinical reasoning can be challenging, and a good assessment program should include multiple assessment methods, cover all components of the clinical reasoning process, and include assessments from both the simulated clinical environment and workplace-based setting [2, 3].

Incorporating OSCEs into assessment programs provides an ideal platform for students to practice and receive feedback on clinical reasoning skills [2, 3]. In particular, the post-encounter note provides opportunities to further probe the clinical reasoning processes of the learner [2, 3]. The post-encounter note was an essential component of clinical reasoning assessment in the United States Medical Licensing Examination (USMLE) Step 2 Clinical Skills (CS) exam [4]. Although more challenging to execute due to resource requirements, post-encounter oral presentations fulfill a similar function [2]. Here, we describe the steps required to implement post-encounter notes in an OSCE program.

V. Schaye (✉) · C. Gillespie
Institute of Innovation in Medical Education, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA
e-mail: verity.schaye@nyulangone.org

R. Crowe
Department of Medicine, New York University Grossman Long Island School of Medicine, New York, NY, USA

C. Beltran
Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, USA

Steps to Implement Post-encounter Notes

OSCEs can be labor- and resource-intensive, and it benefits both the learner and the assessor to maximize opportunities from the encounter—post-encounter notes are one important strategy to achieve this. Other post-encounter experiences, such as oral presentations or computer literature searches related to the scenario, are also opportunities for OSCE enrichment and assessment of clinical reasoning. In this chapter, we will focus on the post-encounter note. There are several key steps to consider when implementing post-encounter notes:

1. Develop a conceptual model of assessment.
2. Create standardized note templates and assessments that map to the conceptual model.
3. Prepare learners by sharing expectations.
4. Recruit and train note assessors.
5. Provide learners feedback and an opportunity for reflection and goal setting.
6. Consider integration of artificial intelligence (AI) into note assessment and feedback.

Step 1: Develop a Conceptual Model of Assessment

At NYU undergraduate medical education, we have adopted three assessment frameworks around which our OSCE assessments are organized: the RIME framework, Entrustable Professional Activities (EPAs), and illness script theory [1, 5, 6]

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_5

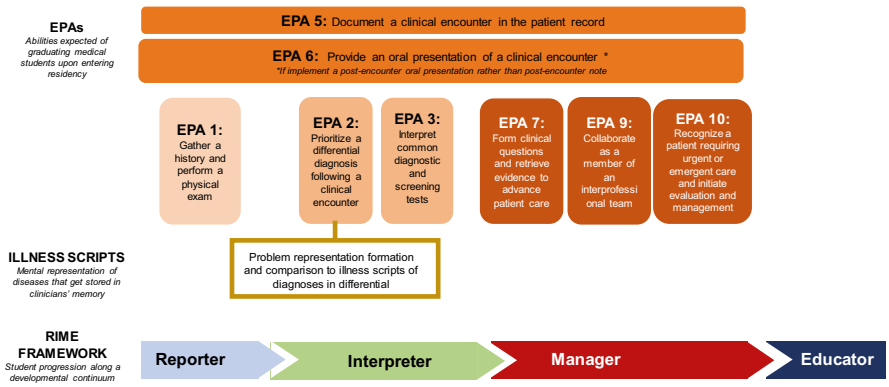


Fig. 5.1 Conceptual model of post-encounter note assessment

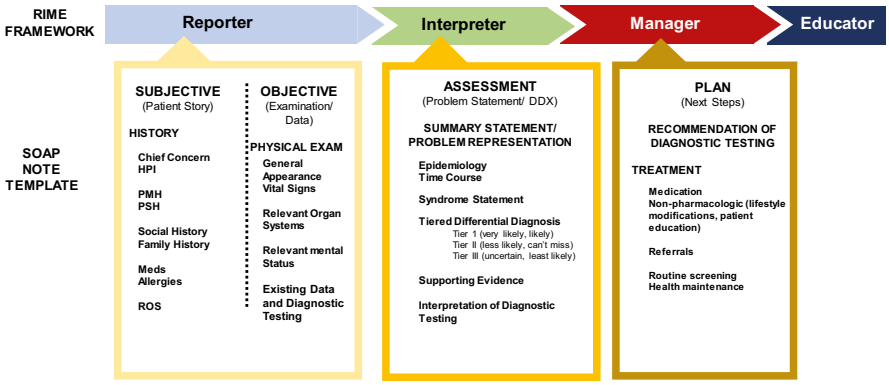


Fig. 5.2 Mapping of note template to assessment framework

(Fig. 5.1). The RIME framework helps describe student progress along a developmental continuum: reporter (trainee can accurately gather and communicate patient information, interpreter (trainee can prioritize and analyze patient problems), manager (trainee can propose reasonable treatment options incorporating patient's circumstances and preferences), and educator (trainee has consistent level of knowledge of current medical evidence and can critically apply that knowledge to patients) [6]. EPAs—the abilities expected of a graduating medical student upon entering residency—offer a broader assessment framework, the majority of which are well captured by performance-based assessment [5]. Lastly, illness scripts are organized mental representations of diseases that get stored in a clinicians' memory and are foundational to teaching and assessing clinical reasoning [7]. Teaching students to form patient summaries as problem representations facilitates retrieval of those illness scripts [1]. These are not the only assessment frameworks to consider, but it is an essential first step to agree upon a shared mental model before implementing a post-encounter note and assessment of clinical reasoning [8].

Step 2: Create Standardized Note Templates and Assessments That Map to the Conceptual Model

Once you have agreed upon a conceptual model, it is essential to create standardized note templates and assessment rubrics that map to the conceptual model (Fig. 5.2). Assessment rubrics should be consistent across the program, allowing data to be accessed and visualized in powerful ways to analyze learner growth and to use for program evaluation purposes. Standardized note templates are useful tools for educators to "build" their post-encounter note tailored to the level of the learner. The note template shown in Fig. 5.3, which includes extensive prompts, details what information should be included in each section of a note to help guide preclinical learners with their note-writing. For advanced clinical learners, there should be little to no prompts as they should be assessed on knowing what information should be included in their note and under which section. As learners develop skills and as OSCE resources allow, additional data can be provided

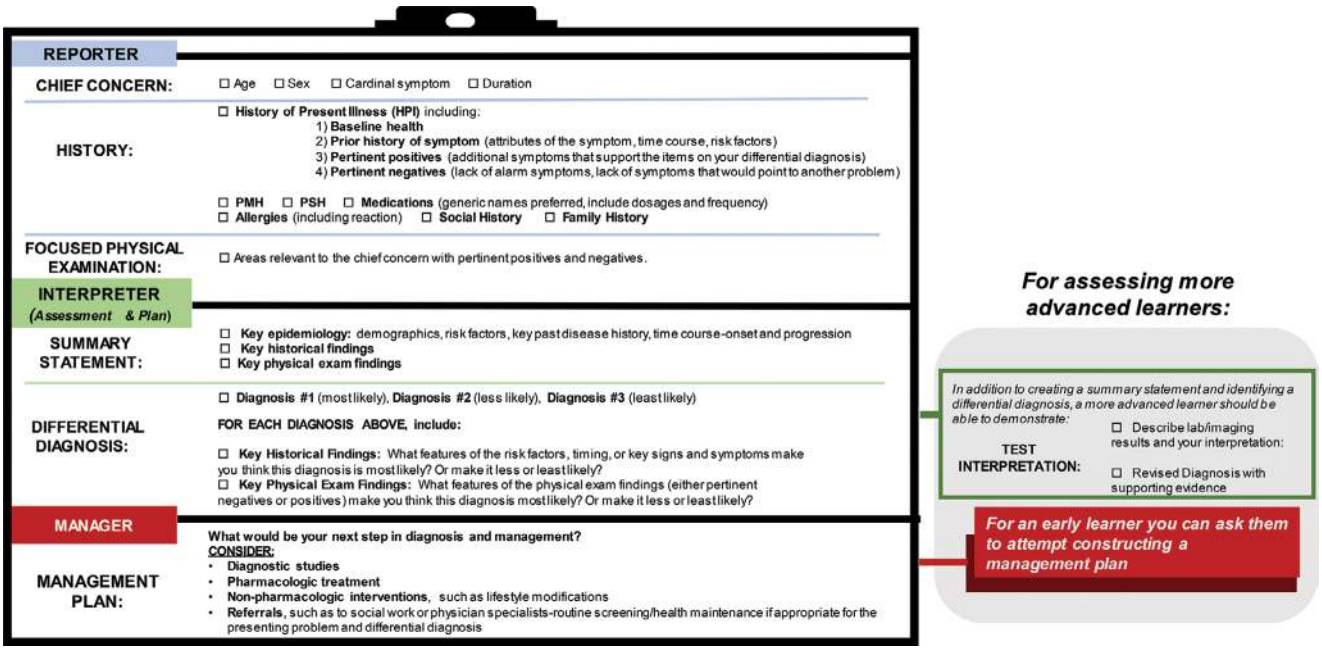


Fig. 5.3 OSCE note template and assessment: preclinical learners receive more extensive prompts, and clinical/advanced learners receive less extensive prompts

to the learner after the patient encounter, such as diagnostic testing (lab results and/or imaging) that would further probe the ability to interpret new data, reconsider the diagnosis, and make recommendations for next steps. Assessment rubric items of these higher-level skills should be added, as highlighted in the call out boxes on the right of Fig. 5.3.

Step 3: Prepare Learners by Sharing Expectations

It is important to provide learners with explicit preparatory instructions and materials with which to prepare for the encounter and post-encounter; clear expectations allow the learner to derive maximum benefit from the session and maximizes the value of the assessment. A consistent shared mental model of assessment and framework for the structure and function of the note are all key steps in setting these expectations. As described above, less prompting is necessary as learners advance.

Step 4: Recruit and Train Note Assessors

Creating clear assessment rubrics (Figs. 5.4 and 5.5) and note assessment scoring sheets (Fig. 5.6) leverages possibilities for note assessment which is always a resource challenge. With clear instructions and processes, the notes can be assessed with minimal training using those available: faculty, admin teams, near-peer senior students, and even peers. Notes should always be de-identified to prevent bias.

Step 5: Provide Learners Feedback and an Opportunity for Reflection and Goal Setting

Consistent note templates and assessment rubrics across the program also facilitates providing consistent feedback to learners. Reflection on post-encounter note feedback is an opportunity to close the learning loop and support master adaptive learning. Analysis of feedback can identify learning gaps through synthesis of self-assessment and external feedback, which encourages goal setting and identification of strategies to achieve those goals. The next OSCE provides practice implementing those skills, which in turn generates iterative feedback with a shared vocabulary and ultimately integrates new skills into daily practice [9]. It can be effective to create a reflection exercise that ensures that the learner processes the feedback provided and does create specific learning goals for future growth. This is also an opportunity to tie reflection back to the conceptual model of assessment. For example, we emphasize the concept of illness script theory and problem representation matching in our post-OSCE reflections (Fig. 5.7).

First Year Practice of Medicine - Note Assessment

REPORTER

1. **Chief Concern (CC) - includes demographics, symptom, time frame**
   - ◯ NOT DONE: Missing at least two of the following: demographics, symptom, time frame OR Not done at all
   - ◯ PARTLY DONE: Missing at least one of the following: demographics, symptom, time frame OR contains extraneous information
   - ◯ WELL DONE: Has all three of the following: demographics, symptom, time frame AND does not contain extraneous information
2. **HPI- baseline health**
   - ◯ NOT DONE: No information is provided regarding patient's baseline health and functioning prior to the presenting problem developing
   - ◯ PARTLY DONE: Insufficient information is provided regarding patient's baseline health and functioning prior to the presenting problem developing to inform the differential diagnosis and management plan
   - ◯ WELL DONE: Complete information is provided regarding patient's baseline health and functioning prior to the presenting problem developing to inform the differential diagnosis and management plan
3. **HPI - Attributes of the symptom includes information concerning:**
   -Onset
   -Character
   -Location (if relevant)
   -Radiation (if relevant)
   -Severity (including impact on baseline function)
   -Duration
   -Timing
   -Alleviating factors
   -Aggravating factors
   -Associated symptoms
   - ◯ NOT DONE: HPI is missing several attributes of the presenting symptom(s) not creating a clear picture of the problem resulting in insufficient information to make an informed prioritized differential diagnosis
   - ◯ PARTLY DONE: HPI includes some of the attributes of the presenting symptoms(s) BUT is missing some key items necessary to provide enough information to make an informed prioritized differential diagnosis
   - ◯ WELL DONE: HPI includes most or all attributes of the presenting symptom(s), creating a clear picture of the problem, providing enough information to make an informed prioritized differential diagnosis
4. **HPI - Risk factors that inform the priortized differential diagnosis are included in the HPI (e.g. smoking history in a patient presenting with cough)**
   - ◯ NOT DONE: Not provided OR missing several key risk factors that inform the prioritized differential diagnosis
   - ◯ PARTLY DONE: Includes some key risk factors BUT misses others that would inform the prioritized differential diagnosis
   - ◯ WELL DONE: Includes most or all key risk factors that inform the prioritized differential diagnosis
5. **HPI - Pertinent negatives that help lower the liklihood of certain diagnoses are included in the HPI**
   - ◯ NOT DONE: Not provided OR missing several key risk factors that inform the prioritized differential diagnosis
   - ◯ PARTLY DONE: HPI includes some key pertinent negatives that inform the prioritized differential diagnosis BUT misses other important pertinent negatives that would inform the differential diagnosis
   - ◯ WELL DONE: HPI includes most or all key pertinent negatives that inform the prioritized differential diagnosis
6. **HPI - Time Course: HPI includes a clear timeline of events in an organized and logical fashion**
   - ◯ NOT DONE: HPI is missing a clear timeline of events
   - ◯ PARTLY DONE: HPI is documented in some chronological order but with lapses in clarity in the sequence of events
   - ◯ WELL DONE: Chronology starts at an appropriate point in the disease timeline, then proceeds logically with minimal breaks in flow to give a clear HPI narrative that allows for clear understanding of the sequence of events
7. **Past Medical History**
   - ◯ NOT DONE: PMH is not documented
   - ◯ PARTLY DONE: PMH is documented BUT is missing key information
   - ◯ WELL DONE: PMH includes most or all key information
8. **Past Surgical History**
   - ◯ NOT DONE: PSH is not documented
   - ◯ PARTLY DONE: PSH is documented BUT is missing key information
   - ◯ WELL DONE: PSH includes most or all key information

Fig. 5.4 Sample note assessment rubric for preclinical learner (without assessment of more advanced skills)

9. Medications are documented including dose, route, & frequency

- ◯ NOT DONE: Medications are not listed OR medications are missing from list
- ◯ PARTLY DONE: Medication list is complete AND includes at least 2 out of the 3 of dose, route, & frequency for each medication
- ◯ WELL DONE: Medication list is complete AND lists dose, route, & frequency for each med

10. Allergies are listed including adverse reaction.

- ◯ NOT DONE: Allergies are not listed OR were not obtained OR NKDA is documented (and clinically inaccurate)
- ◯ PARTLY DONE: Allergy list is incomplete OR adverse reactions are not listed for each allergy
- ◯ WELL DONE: All allergies are listed WITH adverse reactions OR NKDA is documented (if clinically accurate)

11. Social History-includes information about 1) tobacco, alcohol and substance use 2) family and support structure 3) diet and exercise 4) occupation 5) social determinants of health 6) sexual history

- ◯ NOT DONE: Social History is not documented OR extremely limited
- ◯ PARTLY DONE: Social history is included BUT is missing key information
- ◯ WELL DONE: Social history is included and includes most or all key information

12. Family history-includes relevant problems especially those that relate to the presenting concern

- ◯ NOT DONE: Family History is not documented OR extremely limited
- ◯ PARTLY DONE: Family history includes some of the pertinent history BUT is missing key information
- ◯ WELL DONE: Family history includes most or all of the pertinent history

13. General Appearance

- ◯ NOT DONE: General appearance is not documented.
- ◯ PARTLY DONE: General appearance is incomplete OR is not accurate.
- ◯ WELL DONE: General appearance is complete AND accurate.

14. Vital Signs

- ◯ NOT DONE: Vital signs are not documented
- ◯ PARTLY DONE: Vital signs are incomplete OR are interpreted instead of reported
- ◯ WELL DONE: All available vital signs are complete with numeric values AND without interpretation

15. Cardiovascular Exam

- ◯ NOT DONE: no exam documented
- ◯ PARTLY DONE: Exam documented BUT is incomplete OR includes incorrect descriptors of exam findings OR includes subjective information that should be included in history OR includes interpretation
- ◯ WELL DONE: Appropriate focused exam is completely documented AND includes correct descriptors of exam findings AND excludes subjective information that should be included in history AND excludes interpretations

16. Extremeties

- ◯ NOT DONE: no exam documented
- ◯ PARTLY DONE: Exam documented BUT is incomplete OR includes incorrect descriptors of exam findings OR includes subjective information that should be included in history OR includes interpretation
- ◯ WELL DONE: Appropriate focused exam is completely documented AND includes correct descriptors of exam findings AND excludes subjective information that should be included in history AND excludes interpretations

INTERPRETER

17. Summary Statement/Problem Representation: a concise synthesis of the patient's presentation including:
Key demographics, epidemiology and risk factors, temporal pattern of illness, and key signs and symptoms and lab/imaging data (if provided).
Information should be given in abstract terms (2 days->acute), medical terminology (yellowing of skin->jaundice), unified medical concepts (HTN, diabetes and tobacco use->cardiovascular risk factors), and semantic qualifiers (paired opposing descriptors that can be used to compare and contrast diagnostic considerations e.g. monoarticular vs polyarticular).
It SHOULD NOT include diagnoses.

- ◯ NOT provided OR Does not synthesize information (does little more than repeat CC/HPI) OR provides DX/DDX (does not belong here)
- ◯ Begins to synthesize key findings but includes extraneous information OR missing key clinical findings from epidemiology/risk factors, time course, signs/symptoms/labs/diagnostic testing
- ◯ Adequately synthesizes key clinical findings AND includes most key findings from epidemiology/risk factors, time course, signs/symptoms/labs/diagnostic testing BUT does NOT put most key findings in abstract terms, medical terminology, unified medical concepts and semantic qualifiers
- ◯ Adequately synthesizes key clinical findings AND includes most key findings from epidemiology/risk factors, time course, signs/symptoms/labs/diagnostic testing AND puts most key findings in abstract terms, medical terminology, unified medical concepts and semantic qualifiers

Fig. 5.4 (continued)

18. **Diagnosis #1 in the Differential Diagnosis: prioritized differential diagnosis of most likely, less likely, and least likely including "can't miss" diagnoses**
    - ◯ NOT DONE: NOT provided OR Unacceptable diagnosis based on likelihood
    - ◯ PARTLY DONE: Acceptable diagnosis BUT NOT tiered properly (is significantly less likely than designated) OR diagnosis given as category (e.g. malignancy)
    - ◯ WELL DONE: Acceptable diagnosis AND well-tiered
19. **Diagnosis #2 in Differential Diagnosis: prioritized differential diagnosis of most likely, less likely, and least likely including "can't miss" diagnoses**
    - ◯ NOT DONE: NOT provided OR Unacceptable diagnosis based on likelihood
    - ◯ PARTLY DONE: Acceptable diagnosis BUT NOT tiered properly (is significantly less likely than designated) OR diagnosis given as category (e.g. malignancy)
    - ◯ WELL DONE: Acceptable diagnosis AND well-tiered
20. **Diagnosis #3 in Differential Diagnosis: prioritized differential diagnosis of most likely, less likely, and least likely including "can't miss" diagnoses**
    - ◯ NOT DONE: NOT provided OR Unacceptable diagnosis based on likelihood
    - ◯ PARTLY DONE: Acceptable diagnosis BUT NOT tiered properly (is significantly less likely than designated) OR diagnosis given as category (e.g. malignancy)
    - ◯ WELL DONE: Acceptable diagnosis AND well-tiered
21. **Diagnostic Justification for Diagnosis #1: supporting and refuting evidence is given from history, physical and lab/imaging data (if provided). Comparing/contrasting the problem representation of the patient presentation to the illness script of the disease listed in your differential diagnosis.**
    - ◯ NOT DONE: NOT provided OR missing MOST key information (pertinent positives and negatives) from history, physical, and lab/imaging (if provided)
    - ◯ PARTLY DONE: Missing SOME key information (pertinent positives and negatives) from history, physical, and lab/imaging (if provided)
    - ◯ WELL DONE: Includes MOST OR ALL key information (pertinent positives and negatives) from history, physical, and lab/imaging (if provided)
22. **Diagnostic Justification for Diagnosis #2: supporting and refuting evidence is given from history, physical and lab/imaging data (if provided). Comparing/contrasting the problem representation of the patient presentation to the illness script of the disease listed in your differential diagnosis.**
    - ◯ NOT DONE: NOT provided OR missing MOST key information (pertinent positives and negatives) from history, physical, and lab/imaging (if provided)
    - ◯ PARTLY DONE: Missing SOME key information (pertinent positives and negatives) from history, physical, and lab/imaging (if provided)
    - ◯ WELL DONE: Includes MOST OR ALL key information (pertinent positives and negatives) from history, physical, and lab/imaging (if provided)
23. **Diagnostic Justification for Diagnosis #3: supporting and refuting evidence is given from history, physical and lab/imaging data (if provided). Comparing/contrasting the problem representation of the patient presentation to the illness script of the disease listed in your differential diagnosis.**
    - ◯ NOT DONE: NOT provided OR missing MOST key information (pertinent positives and negatives) from history, physical, and lab/imaging (if provided)
    - ◯ PARTLY DONE: Missing SOME key information (pertinent positives and negatives) from history, physical, and lab/imaging (if provided)
    - ◯ WELL DONE: Includes MOST OR ALL key information (pertinent positives and negatives) from history, physical, and lab/imaging (if provided)
24. **EPA 1a. Information Gathering**
    - chief complaint
    - attributes
    - risk factors
    - pertinent negatives
    - ◯ POOR: Neither patient-centered nor adequate to learn about the patient or condition; acquires inaccurate, incomplete or excessive information
    - ◯ BEGINNING: Gathers some key data but lacks organization and completeness; does not use patient-centered interview skills such as open-ended questions, avoiding jargon, active listening, acknowledging emotions, and incorporating responses consistently.
    - ◯ COMPETENT: Obtains organized, accurate and complete history that supports a strong but not fully prioritized differential diagnosis; some missed opportunities for patient-centeredness.
    - ◯ STRONG: Appreciates subtleties and builds rapport by listening to, informing, and involving patients. Uses culturally-sensitive, patient-centered approach to an organized, accurate and complete history supporting a thorough differential diagnosis.
25. **EPA 1b. Physical Examination**
    - complete
    - properly recorded
    - ◯ POOR: Exam is not patient-centered (lacking privacy, draping, comfort, communication) and/or misses major findings and/or uses incorrect technique with insufficient insight and improvement.
    - ◯ BEGINNING: Minimally competent, patient-centered exam; major findings identified but awkward, over-inclusive or incomplete.
    - ◯ COMPETENT: Patient-centered, well-executed examination that is organized, complete, and targeted to relevant areas.
    - ◯ STRONG: Patient-centered, well executed and articulated examination that is organized and complete to task; incorporates history to explore pertinent findings and elicit subtle findings.
26. **EPA 5. Written Documentation**
    - ◯ POOR: Report is disorganized, inaccurate, or incomplete; inadequate medical terminology; does not discriminate relevance; interpretation below expectations
    - ◯ BEGINNING: reports in rote form that relies on norms rather than specific context, lacks supporting detail or interpretation, or is over-inclusive or redundant, with some errors.
    - ◯ COMPETENT: records organized, accurate, prioritized, key information in a timely manner with both comprehensive reporting and synthetic interpretation.
    - ◯ STRONG: records organized, relevant story with sophisticated use of medical terminology for subjective, objective, and interpretive data; concise and complete, analytic approach to understanding disease.

Fig. 5.4 (continued)

27. EPA 2. PRIORITIZING A DIFFERENTIAL DIAGNOSIS. Clinical problem solving. Explains reasoning behind the diagnosis, including epidemiology, risk factors, time course, key signs and symptoms; compares patient representation to common disease illness script

- ◯ POOR: does not identify major clinical diagnoses and does not generate appropriate prioritized differential diagnoses for common conditions.
- ◯ BEGINNING: usually identifies major clinical diagnoses, but includes some unreasonable diagnoses or misses important ones that should be available.
- ◯ COMPETENT: usually identifies major and minor clinical diagnoses and prioritizes them based on likelihood and need for urgent care.
- ◯ STRONG: reliably identifies major and minor clinical diagnoses to generate a prioritized differential diagnosis with evidence from the history, exam, and data that reflect the specific patient.

INTERPRETER

28. Interpretation of Diagnostic Testing - Imaging

- ◯ NOT DONE: NOT provided OR missing most interpretation OR most interpretation is NOT accurate
- ◯ PARTLY DONE: Missing some interpretation OR some interpretation is NOT accurate
- ◯ WELL DONE: Includes most or all interpretation AND most or all interpretation is accurate

29. Revision of Working Diagnosis after diagnostic testing analysis

- ◯ NOT DONE: NOT provided OR Unacceptable diagnosis based on likelihood OR diagnosis given as category (e.g. malignancy)
- ◯ PARTLY DONE: Acceptable diagnosis BUT NOT MOST likely
- ◯ WELL DONE: Acceptable diagnosis AND MOST likely

MANAGER

30. Diagnostic Testing - Directs evaluation toward the prioritized differential diagnosis without recommending unecessary tests that would be low value

- ◯ NOT DONE: NOT provided OR misses MOST key diagnostic tests to evaluate the prioritized differential diagnosis
- ◯ PARTLY DONE: Misses SOME key diagnostic tests to evaluate the prioritized differential diagnosis OR recommends unecessary tests that would be low value care
- ◯ WELL DONE: Includes MOST or ALL key diagnostic tests to evaluate the the prioritized differential diagnosis AND DOES NOT recommend unecessary tests that would be low value care including not recommending any diagnostic tests if clinically appropriate

31. Plan - Treatment - referrals (e.g. procedures, physician consults, inter-professional collaboration)

- ◯ NOT DONE: NOT provided OR misses MOST key referrals to diagnose/treat the prioritized differential diagnosis
- ◯ PARTLY DONE: Misses SOME key referrals to diagnose/treat the prioritized differential diagnosis OR recommends unecessary referrals that would be low value care
- ◯ WELL DONE: Includes MOST or ALL key referrals to diagnose/treat the prioritized differential diagnosis AND DOES NOT recommend unecessary referrals that would be low value care

32 EPA 3. CHOICE AND INTERPRETATION OF DIAGNOSTIC TESTING. EKG, Imaging, Labs

- ◯ POOR: unable to recommend standard data collection, testing, and rationale; poor interpret of basic data; selections put patient at risk.
- ◯ BEGINNING: recommends some data collection and tests for common conditions that is unnecessary and does not consider harm, cost, or resources; inconsistent prioritization and/or interpretation.
- ◯ COMPETENT: recommends data collection and tests for common conditions understanding pre- and post-test probability; diagnostic decisions are consistently reasonable.
- ◯ STRONG: recommends and interprets prioritized, high value data for common conditions, explaining rationale that interrelates complex problems.

Fig. 5.5 Additional items that can be incorporated into the sample note assessment rubric in figure 5.4 in order to assess a clinical learner with more advanced skills

Fig. 5.6 Note assessment scoring sheet

DYSPNEA Patient Note Evaluation Rubric

HISTORY

CC

- ☐ 57M with shortness of breath x 5 mo

Attributes

- ☐ shortness of breath with exertion
- ☐ having to "work hard to breathe,"
- ☐ after climbing one flight of stairs or walking 3 city blocks
- ☐ resolves after a few minutes of rest
- ☐ loss of appetite and thinks he lost some weight because his clothes are looser

Risk Factors

- ☐ smoker
- ☐ asbestos/occ exposure
- ☐ + fam hx CV and pulm dz

Time Course

- ☐ health good, no exercise limits until 5 months ago
- ☐ progressive

Pertinent Negatives

- ☐ denies chest pain, orthopnea, PND,
- ☐ cough, fever
- ☐ leg pain

He also reports after his. His appetite is now back to normal. No sick contacts.

PMH

- ☐ High blood pressure- diagnosed 6 months ago
- ☐ no recent colonoscopy
- ☐ unknown cholesterol and PSA

PSH

wife's death 10 months ago

FH

- ☐ father HTN, sudden death age 76
- ☐ grandchild with asthma, no other pulmonary disease
- ☐ (mother died 74 old age)

SH

- ☐ + construction worker (risk for asbestos exposure),
- ☐ former smoker (1/2-2PPD x 30 yrs, quit 2 yrs ago)
- ☐ 6 drinks EtOH per week, no drugs.

Meds/All

- ☐ HCTZ 25 mg QD
- ☐ NKDA

PHYSICAL EXAMINATION

GA

VS T98.9, HR 74, **RR 18**

- ☐ No lymphadenopathy

CV

- ☐ S1S2 without murmurs, 2+ peripheral pulses bilat, no clubbing, cyanosis or edema

Pulm

- ☐ **clear lungs**, RRR

ASSESSMENT

Summary Statement

- ☐ 57 yo man
- ☐ Smoker, asbestos, fam hx
- ☐ Chronic progressive
- ☐ Exertional dyspnea

DIAGNOSIS #1. COPD	
History Findings	Physical Examination Findings
progressive	afebrile
smoker	Clear lungs
exertional	Slight tachypnea
	(no barrel chest, no clubbing)

DIAGNOSIS #2. Pneumonitis	
History Findings	Physical Examination Findings
smoker	Mild tachypnea
Subacute onset	Normal cardiac exam
No cardiac sxs	

DIAGNOSIS #3. Cardiomyopathy – Ischemic vs. hypertensive vs. IHSS	
History Findings	Physical Examination Findings
HTN	BP high normal on meds
smoker	

DIAGNOSIS #4. Angina	
History Findings	Physical Examination Findings
Vague chest pain	tachypnea
exertional	BP OK on meds
Relieved by rest	Normal cardiac exam
SOB, diaphoresis	Normal lung exam

DIAGNOSIS #5. Lung Cancer	
History Findings	Physical Examination Findings
smoker	Normal lung exam
Progressive dyspnea	Normal cardiac exam
Non prod cough	Normal VS x mild tachypnea
age	

DIAGNOSIS #6. Mesothelioma	
History Findings	Physical Examination Findings
asbestos	Normal lung exam
smoking	Normal cardiac exam
	Normal VS x mild tachypnea

DIAGNOSIS #7. CHF due to cardiomyopathy	
History Findings	Physical Examination Findings
smoker	
htn	
SOB	
exertional	

DIAGNOSTIC STUDIES
Chest X-Ray
EKG
CBC (Anemia)
Chest CT

INTERPRETATION

Imaging:
CXR/Chest CT: right upper lobe mass w/ compression of right bronchus

Dx:
Lung Cancer

Next Steps:
CBC
CMP
CT A/P
Bronchoscopy with biopsy

Fig. 5.6 (continued)

Post-OSCE Guided Reflection and Goal Setting

The objectives of the clerkship OSCEs are to provide an opportunity to demonstrate your communication, physical exam, and clinical reasoning skills in content areas specific to each clerkship discipline. Just like in patient care there is often not a single 'correct' answer on initial presentation but most likely and less/least likely diagnoses that explain the patient's presenting symptoms. Please take this opportunity to reflect on your performance and select at least two of the areas below to generate goals for improvement.

1. **Communication skills, including but not limited to:**
   - Rapport building and cultural humility
   - Information gathering
   - Collaborating on management/patient education

2. **Physical Exam, including but not limited to:**
   - Skill in performing maneuvers
   - Selection of what aspects of the physical exam to perform relevant to the presenting problem
   - Proper documentation of physical exam findings

3. **Problem representation, including but not limited to:**
   - Inclusion of key clinical findings in a concise manor (what are the key pertinent positive or negative demographics, epidemiology, risk factors, temporal pattern of illness, and signs and symptoms that informed your thought process on your prioritized differential diagnosis)
   - Information is given in abstract terms (2 days -> acute), medical terminology (yellowing of skin - > jaundice), unified medical concepts (HTN, diabetes and tobacco use -> cardiovascular risk factors), and semantic qualifiers (paired opposing descriptors that can be used to compare and contrast diagnostic considerations e.g. monoarticular vs polyarticular).
   - Whether a diagnosis is included (should NOT be part of your summary statement)

4. **Differential Diagnosis, including but not limited to:**
   - Compare/contrast the problem representation to the illness scripts of the diagnoses and look up illness scripts of specific diseases as needed to inform your reflection
   - Prioritization of the differential diagnosis according to likelihood (based on prevalence and evidence) and can't miss diagnoses. Are there diagnoses that are more likely than the ones you listed?
   - Supporting the prioritized differential diagnosis with supporting evidence from history, physical exam, and lab/imaging (if available). Were there findings that were expected for the diagnosis that were not present? Are there questions you forgot to ask on history that would inform your prioritized differential?

5. **Plan, including but not limited to:**
   - Recommended diagnostic tests and treatments are informed by the prioritized differential diagnosis and unnecessary tests and treatments are not recommended (high-value care)
   - Look up diagnosis and treatment of specific diseases as needed to inform your reflection
   - Plans incorporate patient's values and preferences and social determinants of health -Referrals integrate interdisciplinary team members when appropriate

Goals for Improvement
The best goals are specific, measurable, attainable, relevant, & time-based (e.g. On the next clerkship OSCE I will be sure to write a problem representation using abstract terms for time course (rather than one day it will be acute):

1. **Goal #1**

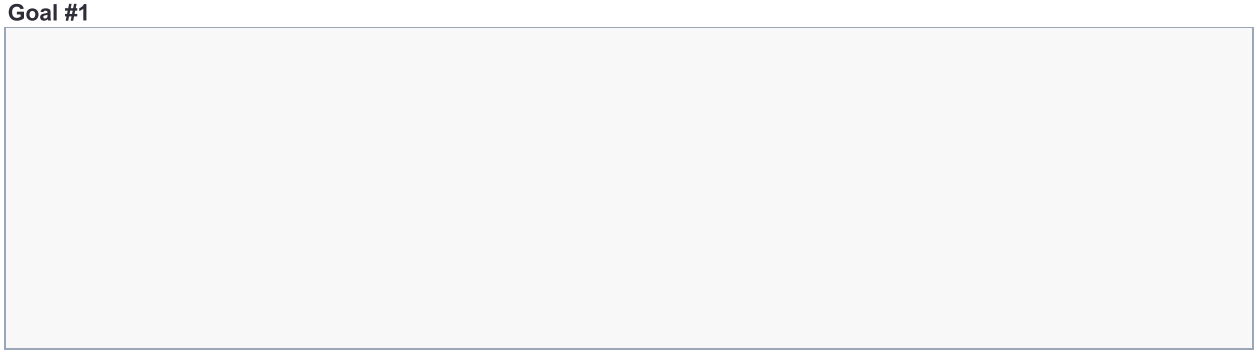


Fig. 5.7 A sample post-OSCE guided reflection and goal setting sheet

2. **Goal #2**

3. **Lastly, we would greatly appreciate your feedback on how to optimize the educational experience of the clerkship OSCEs. Please let us know any feedback you may have:**

Fig. 5.7 (continued)

Goals set by learners can be very rich and incorporate different clinical skill domains.

- On the next clerkship OSCE, I also want to focus on writing a more succinct problem representation. I particularly struggle with grouping symptoms and findings into clusters.
- Become more adept at selecting a focused physical exam based off of my differential diagnosis. For example, checking nails in a case of suspected COPD or lung cancer.
- On the next clerkship OSCE I will ask my patients what they know about a topic before teaching and telling them more information about their diagnosis and potential plan.

Prompting learners to reflect on their prior goals immediately before the next OSCE can help facilitate their implementation.

Remediation

Post-encounter notes that are scored using consistent assessment rubrics provide an opportunity to identify struggling learners and target remediation to the domain of deficiency. Identifying whether the learner shows deficits in information gathering, physical examination, data interpretation, synthesis of key data, formulation of differential diagnosis, or the ability to formulate a treatment plan enables the opportunity to deliver personalized support and remediation plans.

Future Directions: Integration of Artificial Intelligence into Assessment and Feedback

After initial implementation, one of the most ongoing resource-intensive components of post-encounter notes is rater recruitment and training. A potential future direction to overcome this challenge is the use of artificial intelligence (AI)-assisted scoring of post-encounter notes. The NBME had developed a natural language processing (NLP)-based scoring system for Step 2 Clinical Skills notes and planned to supplement expert judgements with computer-assisted scoring [4]. The more recent advances of large language models (LLMs) have greatly enhanced the capabilities of using AI as a tool in post-encounter note assessment and feedback [10]. When integrating AI into assessment, important considerations are whether the OSCE is a formative (low-stakes) or summative (high-stakes) assessment and the performance characteristics of the AI-based assessment tool [11]. As health information technology is such an important component of how we practice medicine, additional future directions should also include integration of the electronic health record (EHR) into the OSCE post-encounter. Ensuring learners are able to use the EHR without compromising the doctor–patient relationship and are proficient in documenting in the EHR are essential skills [12].

References

1. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. NEJM. 2006;355(21):2217–25. https://doi.org/10.1056/NEJMra054782
2. Daniel M, Rencic J, Durning SJ, Holmboe E, Santen SA, Lang V, Ratcliffe T, Gordon D, Heist B, Lubarsky S, Estrada CA. Clinical reasoning assessment methods: a scoping review and practical guidance. Acad Med. 2019;94(6):902–12. https://doi.org/10.1097/ACM.0000000000002618
3. Shen E, Cristiano JA, Ellis LR. The electronic health record objective structured clinical examination station: assessing student competency in patient notes and patient interaction. MedEdPORTAL. 2020;16:10998. https://doi.org/10.15766/mep_2374-8265.10998
4. Pangaro L. A new vocabulary and other innovations for improving descriptive in-training evaluations. Acad Med. 1999;74(11):1203–7. https://doi.org/10.1097/00001888-199911000-00012
5. Core EPA Publications and Presentations. AAMC Website https://www.aamc.org/what-we-do/mission-areas/medical-education/cbme/core-epas/publications. Published May 2014.
6. Lubarsky S, Dory V, Audetat MC, Custers E, Charlin B. Using script theory to cultivate illness script formation and clinical reasoning in health professions education. Can Med Educ J. 2015;6:e61–70. PMCID: PMC4795084 PMID: 27004079
7. Jamieson AR, Holcomb MJ, Dalton TO, et al. Rubrics to prompts: assessing medical student post-encounter notes with AI. NEJM AI. 2024;1(12):AIcs2400631. https://doi.org/10.1056/AIcs2400631
8. Thampy H, Willert E, Ramani S. Assessing clinical reasoning: targeting the higher levels of the pyramid. J Gen Intern Med. 2019;34(8):1631–6. https://doi.org/10.1007/s11606-019-04953-4
9. Cutrer WB, Miller B, Pusic MV, Mejicano G, Mangrulkar RS, Gruppen LD, Hawkins RE, Skochelak SE, Moore DE Jr. Fostering the development of master adaptive learners: a conceptual model to guide skill acquisition in medical education. Acad Med. 2017;92(1):70–5. https://doi.org/10.1097/ACM.0000000000001323
10. Turner L, Hashimoto D, Vasisht, S; Schaye, V. Demystifying AI: current state and future role in medical education assessment. Acad Med. 2024;99(4S):S42–S47. https://doi.org/10.1097/ACM.0000000000005598
11. Young ME, Dory V, Lubarsky S, Thomas A. How different theories of clinical reasoning influence teaching and assessment. Acad Med. 2018;93(9):1415. https://doi.org/10.1097/ACM.0000000000002303
12. Salt J, Harik P, Barone MA. Leveraging natural language processing: toward computer-assisted scoring of patient notes in the USMLE Step 2 clinical skills exam. Acad Med. 2019;94(3):314–6. https://doi.org/10.1097/ACM.0000000000002558

6 Data for Learning and Program Evaluation: Managing, Analyzing, and Reporting OSCE Data

Colleen Gillespie and Tavinder K. Ark

Introduction

Assessing performance is usually a sine qua non—an essential aspect—of the OSCE, whether the results are to be used for formative feedback, summative grading, entrustable or milestone judgments, program evaluation, or all of the above. In Chap. 2 (Organizing OSCEs (and Other SP Exercises) in Ten Steps), Steps 5, 7, and 9 provide broad overviews and practical tips for, respectively, designing assessments, training assessors, and collecting, managing, and analyzing OSCE data. In this chapter, we aim to provide a more in-depth guide for maximizing the educational yield of OSCE data: wrangling, analyzing, visualizing, and reporting on OSCE assessment data to support learning and skill development at all levels, including individual learners, faculty educators and coaches, and education/training programs. While the specifics of data and statistical and psychometric analyses are certainly beyond the scope that can be covered in a chapter, as are the wealth of new programs and applications used in data collection and investigation, we hope to provide readers with a relatively simple and understandable outline for how to approach OSCE data management, reporting, and analysis. Ultimately, we believe that with the explosion of data science advances (e.g., the era of big data, growth and adoption of new analytic programming languages such as R and Python and platforms such as Tableau and other dynamic dashboard applications, artificial intelligence/machine learning, and the increasing analytic sophistication of the education community), there is tremendous opportunity for enhancing our effective and efficient use of OSCE assessment data to achieve the mission of ensuring a highly skilled and competent health professional workforce.

Our sense of the current state of the art in OSCE reporting is that most programs have moved beyond collecting assessment data via paper forms to, at minimum, online data entry/collection platforms and in many cases, sophisticated simulation software systems (such as B-Line/SIM-IQ, EMS, and the like). These approaches have made it possible to have almost instantaneous access to well-structured and complete assessment data—an essential requirement for reporting. However, we believe that many of us are still struggling with how best to turn those data into interpretable, actionable, and educationally meaningful information. Leveraging OSCE data for impact requires, first, clear specification of the goals for and uses of assessment results; then, careful attention to data "pipelines"—the life cycle of data from collection to cleaning to rationalizing to quality assurance to analysis to output; and finally, thoughtful use of the rapidly expanding options for analyzing, reporting, and visualizing data.

Managing Data: Data Processes and Data Pipelines

Data "processing" represents all the ways in which data must be dealt with prior to making use of it—from the point after it has been collected to reporting and interpretation. It includes many aspects with lots of great descriptive terms: data cleaning, data hygiene, data wrangling, data re-structuring, defining data, rationalizing data, etc. We describe in the "Quality Assurance" section below some simple approaches to reviewing the quality of one's OSCE data. Here, we focus on the processes that can be built into a "data pipeline" to maximize the timeliness, interpretability, and usefulness of OSCE data. Automated data capture-to-reporting systems are much more accessible of late and can therefore be leveraged to make the best use of OSCE assessments. These systems accurately collect assessment data in structured, well-defined data fields; provide

C. Gillespie (✉)
Institute for Innovations in Medical Education, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA
e-mail: Colleen.Gillespie@nyulangone.org

T. K. Ark
Data Science Institute, Medical College of Wisconsin, Milwaukee, WI, USA

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_6

opportunities to define the data elements and "map" the items to conceptual domains of assessments (competencies, entrustable professional activities, skill areas); run routine psychometrics to determine the quality of the data and of scores calculated from the raw data; generate scores; provide reports and visualizations of the data to educators, program evaluators, and learners (and for research too); and facilitate interactive queries and analysis of the data via dynamic dashboards.

Metadata: Rationalizing Data

Rationalizing data is the process of ensuring that everyone who works with the data knows all that they need to know about those data elements: information on the context of the data collection, the actual questions and items, the response options and labels, and the identity of the learners and assessors—all the "meta" data (data about data) that defines the assessment data.

Data Dictionaries

The data dictionary is the most basic version of rationalization. Data dictionaries should contain the following: the actual assessment items (verbatim); the variable name or data field name that maps to that item; the labels for the available response options for that item, including both values and labels (e.g., 0 = not done; 1 = partly done; 2 = well done or 1 = Beginning Skills; 2 = Approaching Competence; 3 = Competent; 4 = Able to Teach/Educate); default and permissible values; and any other relevant information (e.g., missing values, values for not sure/didn't observe, not applicable responses, skip patterns). The category or domain of assessment that the item falls into is an essential data dictionary element that then allows "mapping" of items to domains or constructs, as shown in Fig. 6.1, in order to generate summary scores intended to represent performance in a specific competency or skill areas.

Concretely, data dictionaries are built into many data collection, management, and analysis programs but also can be created as a simple spreadsheet (Table 6.1).

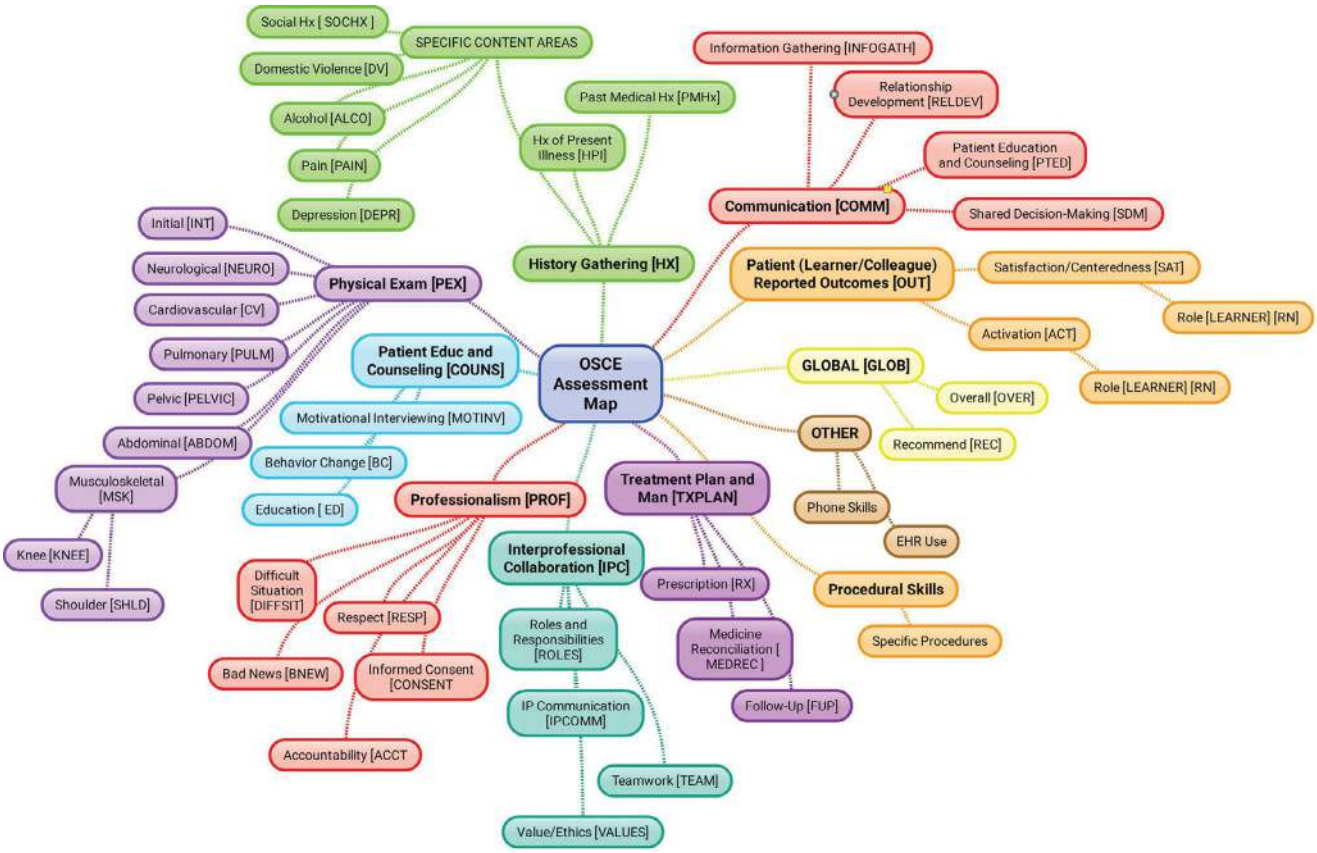


Fig. 6.1 OSCE assessment map

Table 6.1 Sample data dictionary for an OSCE checklist showing overall communication skills and global ratings

Variable/field name	Section header	Field type	Field label	Choices, calculations, OR slider labels
sp_name		Text	SP name	
student_id		Text	Student ID	
student_id_confirm		Text	Please re-enter student ID to confirm	
osce_date		Text		
depr_comm_infogather_elicit	Communication	Radio	Elicited your responses using appropriate questions	0, NOT DONE: Asked leading questions AND asked more than one question at a time | 1, PARTLY DONE: Asked leading questions OR asked more than one question at a time | 2, WELL DONE: Asked one question at a time without leading you in response
depr_comm_infogather_manage		Radio	Managed the narrative flow	0, NOT DONE: Not able to elicit story because questions not organized logically | 1, PARTLY DONE: Elicited main elements BUT illogical order of questions disrupted flow | 2, WELL DONE: Elicited full story by asking questions that facilitated natural flow of story
depr_comm_infogather_clarify		Radio	Clarified information by repeating to make sure s/he understood you on an ongoing basis	0, NOT DONE: Did not clarify/repeat info that you provided | 1, PARTLY DONE: Repeated the info but didn't give chance to indicate whether accurate | 2, WELL DONE: Repeated info and directly invited you to indicate whether accurate
depr_comm_infogather_interrupt		Radio	Allowed you to talk without interrupting	0, NOT DONE: Inappropriately interrupted | 1, PARTLY DONE: Did not interrupt you directly BUT cut your responses short by not giving enough time | 2, WELL DONE: Did not interrupt AND allowed time for you to express your thoughts fully
depr_comm_reldev_concern		Radio	Communicated concern or intention to help	0, NOT DONE: Did not communicate either | 1, PARTLY DONE: Words, actions conveyed intention to help | 2, WELL DONE: Actions, words conveyed intention to help AND concern
depr_comm_reldev_nonverbal		Radio	Nonverbal behavior enriched communication (e.g., eye contact, posture)	0, NOT DONE: Nonverbal was negative | 1, PARTLY DONE: Nonverbal behavior demonstrated attentiveness | 2, WELL DONE: Nonverbal behavior facilitated effective communication
depr_comm_reldev_emotions		Radio	Acknowledged emotions/feelings appropriately	0, NOT DONE: Did not acknowledge emotions/feelings | 1, PARTLY DONE: Acknowledged emotions/feelings | 2, WELL DONE: Acknowledged and responded to emotions in ways that made you feel better
depr_comm_reldev_accepting		Radio	Was accepting (not judgmental)	0, NOT DONE: Made judgmental comments or facial expressions | 1, PARTLY DONE: Did not express judgment but did not demonstrate respect | 2, WELL DONE: Made comments and expressions that demonstrated respect
depr_comm_reldev_jargon		Radio	Used words you (patient) understood and/or explained jargon	0, NOT DONE: Consistently used jargon without explaining | 1, PARTLY DONE: Sometimes used jargon and did not explain | 2, WELL DONE: Explained jargon when used OR avoided jargon completely
depr_comm_pteduc_ask		Radio	ASK: Asked questions to see what you understood	0, NOT DONE: Did not check to see what you understood | 1, PARTLY DONE: Asked if you had any questions BUT did not check understanding | 2, WELL DONE: Assessed understanding by checking in throughout
depr_comm_pteduc_tell		Radio	TELL: Provided clear explanations/ information	0, NOT DONE: Gave confusing or no explanations | 1, PARTLY DONE: Provided some explanation regarding possible diagnosis/treatment but unclear or confusing | 2, WELL DONE: Provided clear explanations; small bits of information at a time and summarized to make sure clear

(continued)

Table 6.1 (continued)

Variable/field name	Section header	Field type	Field label	Choices, calculations, OR slider labels
depr_comm_pteduc_collaborate		Radio	Collaborated with you in identifying possible next steps in diagnosis/treatment	0, NOT DONE: Did not give you opportunity to weigh in on next steps (told you what would happen next) OR didn't discuss next steps at all | 1, PARTLY DONE: Told you next steps and asked then told you patient's view | 2, WELL DONE: Elicited your views on next steps, shared her/his ideas, and mutually developed plan of action
depr_comm_general_comments		Notes	Comments about communication items?	
depr_global_comm_recomm	Global ratings	Radio	Would you recommend this medical student to a friend or family member for his/her overall communication skills?	0, NOT recommend | 1, recommend with RESERVATIONS | 2, recommend | 3, HIGHLY recommend
depr_comm_recomm_comments		Notes	Please explain your recommendation choice ESPECIALLY if did NOT RECOMMEND or RECOMMENDED with RESERVATIONS:	
depr_global_profnl		Radio	Overall, how would you rate this medical student's professionalism?	0, not at all professional | 1, somewhat professional | 2, mostly professional | 3, completely professional
depr_profnl_comments		Notes	Please explain your professionalism rating	
depr_general_comments		Notes	Any other overall comments you think it's important to share:	

A "mapping" stage can be built into the data pipeline to map existing assessment items (questions and responses) to relevant conceptual frameworks including foundational competencies such as communication, history gathering, physical examination, clinical reasoning, and competency sets such as the Physician Competency Reference Set (PCRS) or the ACGME Competencies and Entrustable Professional Activities such as the AAMC's core set of EPAs expected of day one residents. This provides flexibility in how raw data are "scored" and conceptualized in terms of adjusting for changes in prevailing models and the goals and functions of the assessments but also in terms of being responsive to underlying data quality. Rather than automatically calculating a summary score, this approach allows for initial evaluation of the reliability and validity of that score based on the specific sample (of data, learners, administration of the OSCE, etc.) and then determination of whether the "score" meets minimum quality requirements. This builds quality assurance into the process of defining categories and calculating category scores and, therefore, can substantially increase rigor, reproducibility, and, ultimately, trust in the results.

Case Database

The complexity of OSCEs requires additional levels of data definition/rationalization that are critical for understanding and interpreting the results, including definitions or information on the following:

- Raters (often standardized patients but can include other standardized roles and external observers)
- The OSCE itself (type of OSCE in terms of the goals of the OSCE and the function of the OSCE in the curriculum, day/time, order of stations)
- the cases/stations (tasks asked of the learner, information available to the learner, nature, and characteristics of the case)

There is a wealth of information contained in OSCE case scenarios and instructions—all of which are important contextual influences and ideally should be captured, to be available for inclusion in the analysis and interpretation of assessment results.

For these reasons, we strongly recommend that programs maintain a database of OSCE cases. This not only keeps track of what the "examination" in the OSCE consisted of (the clinical conditions, the patient characteristics, the type of tasks required of the learner) but also facilitates use, reuse, continuous refinement, and improvement of OSCE cases. Good OSCE cases, as described elsewhere in this manual, are rich in description, provide a fully developed character and history for the patient, and, if used repeatedly, can be updated and titrated to achieve the intended objectives. If key elements of the cases are coded and included with the assessment data, programs can not only describe learners' clinical experience and exposure across OSCEs in a program of assessment but also explore patterns of performance based on clinical conditions or types of cases. This

information is helpful at the **individual learner** level: does a learner's performance in an OSCE with multiple stations or across a program of assessment (multiple OSCEs) show strengths and weaknesses according to clinical features of the case (cardiovascular vs. renal), demographics of the patient (adolescent vs. geriatric), setting (primary care vs. acute), or task (focus on diagnosis vs. focus on management)?

In addition, these same questions can be answered at the **program level** and are therefore particularly useful for evaluating educational programs and curricula. If cases are defined, educators can describe the full clinical experience of learners across a program of OSCE assessment and integrate this with their clinical experience with real patients. This encourages thoughtful use of OSCEs to fill gaps in clinical experience for rare conditions or different patient populations and also provides a comprehensive view of learners' clinical experience/exposure.

Thus, the description of OSCE cases can be part of an educational program's curriculum map and in turn used to evaluate achievement of program objectives. Is a cohort of learners able to accurately assess substance use risks? Has the curriculum adequately prepared learners to perform a neurological examination? Do incoming students have foundational behavior change counseling skills, or should additional training and skills development be incorporated into their curriculum?

Finally, coding of cases allows for better understanding of the case specificity of core skills and competencies—when are learners' skills transferable across different clinical, setting, and patient contexts (and when they are not transferable)? To which contexts do skills generalize?

NYU Grossman School of Medicine (NYUGSOM) has a case bank of over 700 SP cases and evaluation materials which have been used across the continuum (undergraduate, graduate, and continuing medical education) over the past 22 years (starting in 2001). Our cases have been used by over 30 different programs for performance-based assessment, teaching, and the evaluation of programs, educational innovations, and interventions while providing learners with the opportunities to gain practice and receive feedback.

We have developed cases around the following content areas:

- **Core (common) medical cases** (learner-standardized patient cases) focused on evaluating clinical skills around assessment, physical exam, diagnosis, treatment, and management of various clinical problems that range from common to rare and chronic to acute conditions.
- **Teaching/precepting cases** (learner-standardized trainee cases) focused on assessing skills in teaching or precepting.
- **Inter-professional cases** that assess learners' skills in collaborating with other healthcare providers (e.g., standardized nurses or physician assistants).
- **Structural competence, social determinants of health, and health disparities/health equity cases** that assess either (1) learners' skills in caring for a diverse population of patients across different ages, religions, races/ethnicities, gender identities, etc., (2) skills on how to deal with situations of bias or discrimination, or (3) skills on how to address social determinants of health and their impact on patients' health.
- **Professionalism cases** that assess learners' skills in dealing with difficult situations such as breaking bad news, obtaining informed consent, and disclosing and managing medical errors.
- **Targeted cases:** Designed to assess specific challenges, for example, telehealth, rare or emerging issues (e.g., newly available genetic counseling or public health emergencies such as MPox or dengue fever), new evidence-based practice guidelines, and patient populations/conditions not routinely part of a learners' clinical training (e.g., rural patient care in urban primary care residency programs or transgender patients in clinical systems where their numbers limit the likelihood of a learner having an opportunity to treat).

When creating this case database, we coded cases based on a number of categories: level of the learner, characteristics of the SP clinical conditions and symptoms, visit type, goal of the case, skills assessed, social information, medical history, etc. See Table 6.2 for an example of our case database data dictionary.

Data Pipelines

The huge growth in data and commensurate appreciation for the power of data science and analytics have been made possible by (and have led to expansion in) the availability of technologies for working with data. These range from programs for capturing data, like the software systems that many simulation centers use to "run" OSCEs and simulations and to assess performance of learners, to sophisticated, dynamic dashboards that make OSCE performance results available to learners and educators in near real time. What connects these programs are data pipelines—structured means for moving data throughout the stages of use, from data processing and quality assurance to score computations, analysis, visualization, and reporting.

Data can be moved manually—exporting a data file and then importing into a specific application. This is considered manual because one has to download the data file every time someone makes a change in the data capture system. However, it is much more efficient to create linkages between

Table 6.2 Case database data dictionary

Field name	Field type	Field label	Response options	Notes
case_name	Text	Case name		
case_year	####	Case year		Version control
case_author	Text	Case author		
sp_instructions_file_upload	File	SP instructions		Store with data
learner_instructions_file_upload	File	Learner instructions		Store with data
assessment_file_upload	File	Assessment		Store with data
chief_complaint	Text	Chief complaint		Describe
medical_conditions	Text	Medical conditions		Describe
case_type	Choices	Case type	1. Clinical (provider) 2. Clinical (team/IPC) 3. Teaching/Precepting 4. Other	
visit_type	Choices	Visit type	1. New 2. Continuing 3. Other	
visit_acuity	Choices	Visit acuity	1. Acute 2. Chronic 3. Other	
vite_visit	Choices	Visit setting	1. Inpatient 2. Emergency Dept 3. Outpatient 4. Walk-in/urgent care 5. Other	
visit_modality	Choices	Modality of visit	1. In-person 2. Telephone 3. Video	
patient_goal	Text	Patient goal		Underlying goal of the patient
learner_tasks	Text	Tasks expected of learner		Describe what learner is expected to do
SP_age	##	Patient age		
SP_race_ethnicity	Choices	Patient race/ethnicity	Relevant choice options	
SP_gender_identity	Choices	Patient gender identity	Relevant choice options	
…….				
core_case_functions	Choices (select all)	Main functions of the case	1. Conduct comprehensive history 2. Conduct focused history 3. Screening 4. Targeted physical exam 5. Well visit/"physical" ….	

data collection, analysis, and reporting programs so that data are automatically "piped" throughout these processes. To establish such mechanisms, it is important to ensure the software you are working with has an application programming interface (API). These are protocols that allow different software applications to communicate and interact with each other, providing a standardized way for developers to access data appropriately.

Many working with OSCE data use manual processes and don't have the luxury of data pipelines. For example, we might collect data in a structured data collection program like REDCap [1], Qualtrics, or Google Forms. Those data collection platforms are great for capturing data consistently and accurately, using built-in data fields with controlled options for responses, automated skip patterns, and enforcement of required responses. However, these programs are not designed to conduct sophisticated analyses or create summary reports with unique permission access. Therefore, we often have to export the data for use in other more analytically focused programs. Similarly, while software systems developed for simulation centers like B-Line (SIM/IQ) or EMS are very effective in collecting the data, their ability to analyze and report on the data is limited to a single approach that cannot be customized to the needs of each OSCE and the needs of the learners, programs, and institutions. Therefore, we recommend using statistical programs or languages (e.g., SPSS, R, Python) for customized analysis and reporting of OSCE assessments. These programs provide tremendous

flexibility and can therefore be harnessed to answer almost any question asked of OSCE data.

The growth in data has also led to the development of "dashboard" programs—ways of displaying data in one place (often with little scrolling) so that insights from the data can be generated in near real time. These dashboards give users the paradigm-shifting capability to be able to manipulate and interact with the data through filtering, choosing ways of visualizing data, and comparing data across samples. See Fig. 6.2 for a brief description of the features and limitations of these ways of analyzing and reporting on data. These resources, taken together, provide almost unlimited capabilities for making sense of our data so that it can be used for learning, assessment, and program evaluation.

The ideal approach to leveraging these resources goes beyond manually moving data from application to application and, instead, to a fully integrated and "automated" data pipeline (see Fig. 6.2 for a visualization). Conceptually, the idea is that data are seamlessly "processed" and visualizations/reports generated rapidly for insight, action, and decision-making. Imagine quality assurance analyses that can be automatically and immediately generated for every administration of an OSCE (or even partly through to provide mid-course corrections). Imagine educators that can view sophisticated analyses of performance in near real time and therefore are able to incorporate those analyses into post-OSCE debriefing sessions. Imagine learners, at the end of their OSCE, being able to access their results at multiple levels: granular item responses for learning and improvement, scores compared to benchmarks to understand readiness for upcoming developmental stages, and analyzation of their own data in terms of strengths and weaknesses through interactive visualizations that encourage and support them in self-directed learning and deliberate practice.

Examples of OSCE data pipelines in practice: Below, we describe two examples of OSCE data systems at two medical schools—the first more broadly and the second with much more specificity, reflecting the backgrounds of the authors (one a generalist social scientist/medical education researcher and the other an education researcher with a comprehensive background in psychometrics and data science).

NYUGSOM OSCE Data Pipeline

At NYUGSOM, our education data warehouse serves as the center of a rich data ecosystem (closely aligned with our clinical data systems) that includes OSCE data as an essential element. B-Line/SIM-IQ software in our Simulation Center is used to collect OSCE assessment data. SP checklist reports (completed versions of the assessment forms) are almost immediately available to learners through the B-Line portal. However, calculation of scores based on quality assurance and psychometric analyses are conducted through a combination of dynamic dashboards (Tableau) and statistical programming languages (largely R).

Data from B-Line/SIM-IQ is downloaded directly to our Education Data Warehouse where it can then be processed and queried for many different purposes. OSCE data are passed through a mapping interface that links individual questions to specific domains of competence or EPAs to calculate scores. Once mapped, these data are pulled into Tableau dashboards and html-based interactive views created through R and one of its reporting applications, Shiny App. OSCE administrators can review the quality of the data, analyze internal consistency of categories, see distributions and rater effects, and refine calculations and standards via the Shiny App. OSCE administrators, educators, and learners can view OSCE data through the Tableau dashboard and analyze the data by choosing cohorts, OSCEs, cases, categories of competence, assessment types (clinical skills vs. written notes), level of analysis (items vs. scores), and time/sequence (longitudinal views of performance and/or comparisons over time in relationship to curricular changes) for exploration and comparison.

Lastly, customized reports are generated using R and Shiny App to provide learners with constructive and targeted feedback, educators with information on the performance of a defined group of learners, and the program evaluation team with an aggregate view of learner competencies to evaluate achievement of educational program objectives.

Code for data processes, quality assurance, scoring, and visualization/reporting is stored in GitLab (a local version of GitHub, a code repository) to facilitate reproducibility and re-use. The code can be re-used with almost any set of OSCE data to provide a consistent set of meaningful analyses, reports, and visualization as well as a standard way of cleaning, wrangling, and rationalizing the OSCE data.

Medical College of Wisconsin (MCW) OSCE Data Pipeline

The pipeline at MCW uses data intake/capture systems, database, and data reporting within a model, view, and controller concept. Data are ingested from a wide variety of sources and stored in Postgres (or PostgreSQL, a relational database management system). Each database has a metadata dictionary that is used as the controller of input to output for reporting. The data is reported out in a variety of mechanisms using R and Python into dashboards, PDF reports, and web pages.

MCW has developed a sophisticated data pipeline that leverages the principles of the model-view-controller (MVC) architecture (see Table 6.3) to streamline data intake, storage, and reporting processes. The architecture and components of the MCW data pipeline integrates various technologies to provide robust and flexible data reporting solutions and the capacity to stay up-to-date as technology changes with plug-and-play-like tools.

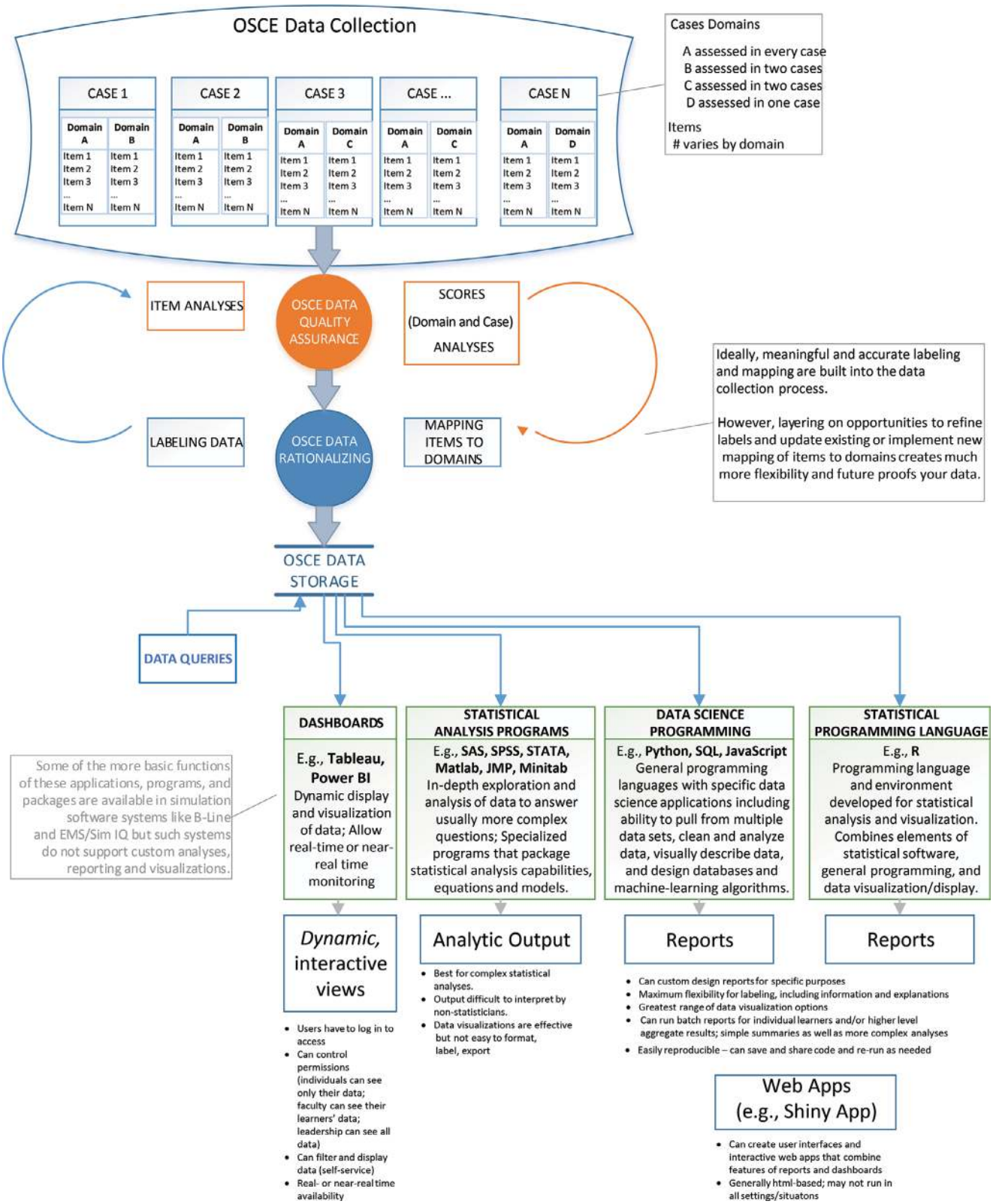


Fig. 6.2 Data processes and pipelines

Table 6.3 The model-view-controller (MVC) concept: the MVC architecture is a software design pattern that separates an application into three interconnected components

Model	Represents the data and business logic of the application. In the MCW data pipeline, the model component is embodied by the PostgreSQL databases and the metadata dictionary. The model handles data storage, retrieval, and integrity, serving as the foundation for all subsequent operations
View	Represents the user interface
controller	Acts as an intermediary between the model and the view, managing user inputs and updating the model and view as necessary. At MCW, the metadata dictionary functions as the controller, guiding the flow of data from intake to storage to reporting. The controller ensures that data transformations and calculations are consistently applied, enabling accurate and reliable reporting

Data Intake and Capture Systems: The first step in the MCW data pipeline is the data intake process, which involves capturing data from a multitude of sources. These sources can include electronic health records (EHRs), data forms (e.g., Qualtrics or REDCap), OSCE systems, and other external databases. The intake systems are designed to handle structured, semi-structured, and unstructured data, ensuring comprehensive data capture. To facilitate seamless data ingestion, MCW employs a combination of custom-built data capture tools and commercial software solutions. These tools are capable of performing real-time data extraction, transformation, and loading (ETL) operations, ensuring that data is cleaned, standardized, and ready for storage.

Data Storage with Database Software: Once ingested, data is stored in Postgres databases. Databases should be chosen for their robustness, scalability, and support for complex queries and transactions. The databases are designed with a focus on maintaining data integrity and security, employing techniques such as data encryption, access controls, and regular audits. The database at MCW is accompanied by a comprehensive metadata dictionary. This dictionary acts as the controller within the MVC architecture, managing the relationships between data elements and ensuring consistent data definitions and structures across the system. The metadata dictionary also plays a crucial role in data governance, providing documentation and lineage information that supports data quality and compliance efforts.

Data Reporting Mechanisms: MCW utilizes a range of tools to generate reports and visualizations from the stored data with R and Python. We create interactive dashboards using tools such as Shiny (R), Dash (Python), and React (JavaScript) to provide real-time data insights and allow users to explore the data through interactive visualizations. Regardless of dashboards, many users still want to be able to print data and reports; thus, automated PDF reports are generated using packages like R Markdown and ReportLab (Python) or a user-friendly print feature of html. These reports compile data analyses and visualizations into a format that is easily shareable and accessible offline. Data is also presented on web pages, leveraging frameworks like Flask (Python) and R Shiny Server. These web pages provide an accessible platform for broader dissemination of data insights to stakeholders across the institution.

The data pipeline at MCW exemplifies the effective application of the MVC architecture in a complex data environment. By integrating advanced data intake systems, robust Postgres storage solutions, and powerful reporting tools, MCW has created a scalable and flexible infrastructure that meets the diverse needs of its users. This architecture not only enhances data management and reporting capabilities but also ensures data integrity, security, and compliance, supporting MCW's mission of advancing health through research, education, and patient care.

GitHub is central to managing these data pipelines and processes as it offers a robust platform for developing, maintaining, and deploying applications focused on data management, such as the data pipeline architecture at MCW. By adopting best practices for branching, merging, and security, the MCW data pipeline projects remain well-organized, secure, and maintainable. This integration simplifies the development process, enabling the team to concentrate on creating and maintaining a robust data management system.

Quality Assurance of OSCE Assessment Data

Overview of Quality Assurance

The first goal of analyzing OSCE data should always be to begin to take stock of the quality of the assessment data. Without evidence of its accuracy, we cannot interpret or act on the results. Thus, we must ensure all values are encoded correctly (e.g., 0 = not done) and that all items intended to measure a construct were present and accurately represented on the checklists. Once that can be verified, the next step is to analyze the data for its reliability and validity. Psychometrics is broadly defined as the scientific study of measurement that began in psychology and sought to determine how closely observations (measures) match theoretical constructs (latent variables) [2]. This helps us gauge the meaning of OSCE data by exploring the degree to which assessments are consistent (reliable) and capture intended constructs such as communication skills and physical examination competence and entrustability in clinical reasoning (validity).

There is some debate as to whether OSCE assessments can routinely and cost-effectively achieve the high standards of rigor required by the most formal applications of psychometrics and classical test theory [3, 4]. However, more

recent validity frameworks (e.g., Messick's [5] approach of integrating the usefulness of assessments—consequential validity—into empirical examinations of the trustworthiness of measures) emphasize the holistic role of OSCEs in a program of assessment and highlight the value these performance-based data provide to learners, educators, and educational programs. Practical guidance for increasing the quality of assessments generally includes expanding the number of stations, reducing the number of raters (one assessor per station whenever possible), providing intensive training and feedback to raters, and continuously refining the assessment items and options (see [6–8], among others).

At its most essential, the question of measurement quality in OSCE assessments is the degree to which the evidence and rationale support inferences and actions based on the OSCE test scores—that it supports the goals of the OSCE. To that end, we focus on working through the various goals of OSCEs and the simple analytic approaches available to non-psychometricians in beginning to establish the quality of the OSCE data. For those who wish to dig into psychometrics more deeply, we suggest referring to helpful overviews (e.g., [9, 10]) as well as the literature on generalizability theory and latent profile analysis and consultation with experts in these areas for further depth.

Quality Assurance Reporting and Visualizations

We start by describing a simple set of steps (and examples of reports and visualizations) for initial quality screening of OSCE assessments, beginning at the level of individual items and then moving to the level of score or scale development, wherein domains are usually represented by summarizing performance across a number of items intended to represent specific areas of competence or skill.

Item-Level Analyses

1. Examine the frequency of missing data. Note that missingness can be virtually eliminated in most data collection programs/applications by requiring completion of the item to continue on to the next. If data are missing, analyze patterns to determine why. Methods for imputing missing values are available but can be complex and depend on certain requirements being met.
2. Review the data for permissible values (values must be within specified ranges on the assessment form—usually can be ensured by not allowing data entry to continue if the entered value is not valid, i.e., is not one of the available options).
3. Analyze the frequency of responses for each item to describe variation and ensure they are what is to be expected based on the item and OSCE goals for that level of learner. For example, if all students are scoring perfectly on an item, this may not be an anomaly as they may be expected to get that item correct based on their level of development. However, if all responses for an item are at the lowest or the highest level, this may indicate that the item is not aligned well to the developmental level of the learners and may not provide the ability to differentiate among learners. If the goal of the assessment is to determine mastery or set thresholds for competence, the lack of variation may reflect the intended goal of documenting achievement of required levels of performance—that is, that almost all of the learners have achieved the intended benchmark. A simple table of the distribution of responses for all learners is a useful way of reviewing these data (Fig. 6.3).
4. Examine the frequency of responses for individual items by assessors (to identify rater bias, consistencies, or patterns), seen in Table 6.4, and across cases/stations (to identify case effects—the performance of an item or the assessment of that item may be influenced by the case within which it occurs), shown in Table 6.5. Other effects can be explored, such as session, time of day, or learner level, in a similar manner. Cross tab analyses (response options crossed with assessors or cases) are an easy way to explore these data and uncover patterns by raters or cases.

   Note: See score-level analyses below for how to explore these same effects at the score rather than item level.

Score-Level Analyses

1. One item is generally not an accurate representation of a student's skills or competency in a domain. Thus, items are often summated or averaged to derive a "domain" score. Domains of performance in an OSCE are usually represented by summarizing performance across a number of items intended to represent a specific area of competence or skill—that summary is commonly known as deriving a "score." Scores are usually based on empirically supported a priori evidence of concepts or constructs—underlying skills, attributes, or practices that reflect a distinct category of performance using analyses such as factor analytic modeling. Several types of analyses should be conducted prior to finalizing the calculation of scores. It is important to note that the numeric structure (continuous vs. ordinal) of the scores must be taken into consideration when choosing an appropriate statistical analysis. For example, while we describe the use of Cronbach's alpha for estimating internal consistency below, that approach is appropriate for continuous but not ordinal (numerical categories) data. Information below is meant to provide more of a conceptual framework than a specific "how-to"—we direct the reader to more specific references for actual application.

Case Summary: The learner is tasked with "giving sign-out" to an intern (SP) coming on-service with a reputation for being disorganized and a loner. The intern is scripted to make clear comments about how difficult residency is and admits to being stressed. Case background includes being new to the city, stressed about residency and meeting criteria for major depression and risky alcohol use.

This station is designed to assess the resident's ability to:

- Recognize substance abuse in a colleague
- Recognize depression in a colleague
- Appropriately discuss these issues with the colleague

Case Specific Skills

Screen for Depression	**N**	**Not Done**	**Partly Done**	**Well Done**
Checked-in with you about your emotional state e.g. "Do I need to worry about you?"		Avoided talking about anything personal, did not ask you if he/she needed to worry about you	Acknowledged your emotional state but did not express concern	Checked in and was willing to talk about your issue
Residents	25	8%	4%	88%
Screened for **depression**: 1. Do you experience a depressed mood most of the day, particularly in the morning? 2. On a daily basis, do you no longer enjoy activities you normally did?		Did not ask about depression	Asked only partially/generally about depression (e.g., "How are you feeling?")	Learned about both depressed mood and daily enjoyment of activities
Residents	25	36%	60%	4%
Asked specifically about other **symptoms**: Weight loss or gain; Appetite change; Hyperactive or slowed down in movement; Loss of energy (fatigue); Feelings of worthlessness or guilt		Did not ask for any additional information	Asked about only 1-2 symptoms	Asked about AT LEAST 3 symptoms
Residents	25	96%	4%	0%
Asked about **suicidal ideation**		Did not ask about suicide	Asked if you had thoughts of hurting yourself	Asked specifically about suicidal ideation
Residents	25	84%	16%	0%
Assessed Current Life Situation				
Assessed for **recent stressors** (marital discord and/or family problems, financial problems, etc.)		Did not ask about any recent stressors	Asked about recent changes AND/OR stressors but did not explore in any detail	Asked in detail
Residents	25	80%	12%	8%
Determined **support system**		Did not ask about support system	Asked general question about support system but did not explore ("have you ever told anyone this?")	Attempted to identify a person with whom you could discuss your situation (within your own family OR through the hospital)
Residents	25	12%	64%	24%
Substance Use				
Assessed **alcohol use**		Did not ask any questions about substance use	Asked about substance use but DID NOT quantify alcohol use	Asked about substance use AND quantified alcohol use (# drinks per week, #

Fig. 6.3 Distribution of responses (in table and histogram format): impaired colleague case-specific skills summary

				of drinks/drinking night)
Residents	25	56%	40%	4%
Screened for **alcohol abuse**:		Asked none of these questions	Asked 1-2 of these questions	Asked 3 or more of these questions
Residents	25	88%	12%	0%
Follow-up				
Personally mitigated feelings of **aloneness or isolation** (“*We* will…” or “*I*’m available”)		Does not attempt to personally make you feel less isolated/alone	Suggests alternative sources of support but does not personally offer him/herself	Personally includes him/herself in your support system
Residents	25	20%	20%	60%
Directed you to seek **help for depression**		Did not direct you to help for your depression	Suggested you seek help (e.g., “Maybe you should try getting some help.”)	Directed you towards specific resources for your depression
Residents	25	0%	28%	72%
Directed you to seek **help for substance abuse**		Did not direct you to help for your substance abuse	Suggested you seek help (e.g., “Maybe you should try getting some help for that.”)	Directed you towards specific resources for your substance abuse
Residents	25	4%	28%	68%

Histogram depiction of Screen for Depression items

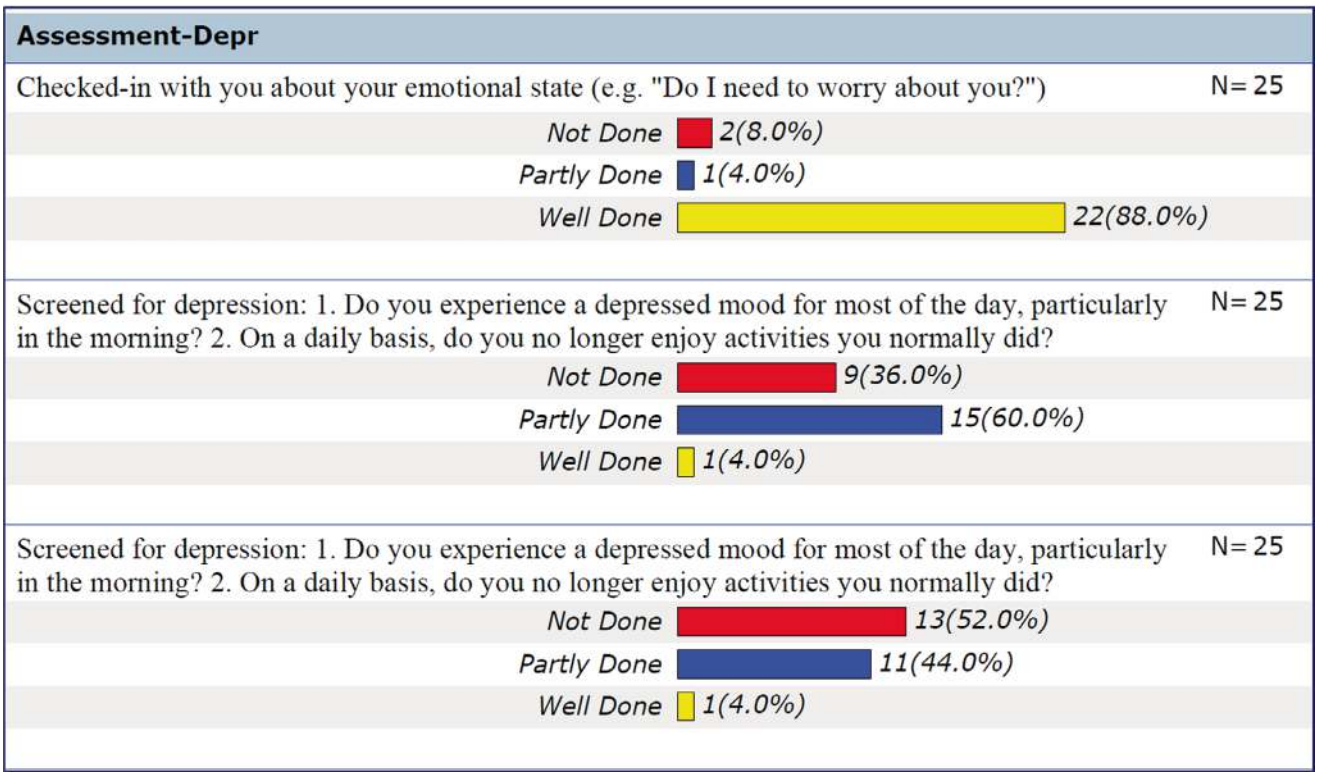


Fig. 6.3 (continued)

Table 6.4 Distribution of responses by rater (N = 40 learners; 20/SP)

Communication skills: information gathering				*Comments*
Started with open-ended questions and used appropriately throughout	**Not done**	**Partly done**	**Well done**	**SP 1 showed no variation within this item, while SP 2 did**
	Started with closed, yes/no questions	Began with open-ended questions but stopped prematurely	Started with open-ended questions and continued using them as appropriate	
SP 1	**0**	**0**	**20**	
SP 2	**5**	**10**	**5**	
Total	**5**	**10**	**25**	
Asked what you thought caused your symptoms/ underlying concerns	**Not done**	**Partly done**	**Well done**	**SP 1 continues to show no variation (now between items), while SP does vary**
	Did not ask	Asked but did give opportunity to fully share thoughts/concerns	Asked so that could fully share thoughts and concerns	
SP 1	**0**	**0**	**20**	
SP 2	**0**	**10**	**10**	
Total	**0**	**10**	**30**	
Managed the narrative flow of your story	**Not done**	**Partly done**	**Well done**	**SP 1 using two scale options for this item but still hasn't given a "not done" rating**
	Not able to elicit narrative because questions not organized logically	Elicited main elements of narrative, but illogical order of questions, leading questions or multiple questions disrupted flow	Elicited full narrative by asking questions that facilitated natural flow of story	
SP 1	**0**	**5**	**15**	
SP 2	**6**	**12**	**2**	
Total	**6**	**17**	**17**	

SP 1 and 2 differ in distribution. SP 1 s' ratings show little variation ("dove"); assuming sample of learners for each SP are similar (e.g., determined simply by random schedule), review of SP assessments warranted

Table 6.5 Distribution of responses by case (*N* = 40 learners)

Communication skills: information gathering				*Comments*
Started with open-ended questions and used appropriately throughout	**Not done**	**Partly done**	**Well done**	***Similar distribution across two different cases: These skills likely to be consistent across cases***
	Started with closed, yes/no questions	Began with open-ended questions but stopped prematurely	Started with open-ended questions and continued using them as appropriate	
Case 1	**10**	**15**	**15**	
Case 2	**10**	**12**	**18**	
Asked what you thought caused your symptoms/ underlying concerns	**Not done**	**Partly done**	**Well done**	
	Did not ask	Asked but did give opportunity to fully share thoughts/concerns	Asked so that could fully share thoughts and concerns	
Case 1	**7**	**13**	**20**	
Case 2	**5**	**9**	**26**	
Managed the narrative flow of your story	**Not done**	**Partly done**	**Well done**	***Case 1 may be more complex and therefore learners struggle to manage the narrative flow***
	Not able to elicit narrative because questions not organized logically	Elicited main elements of narrative, but illogical order of questions, leading questions or multiple questions disrupted flow	Elicited full narrative by asking questions that facilitated natural flow of story	
Case 1	**15**	**15**	**10**	
Case 2	**3**	**20**	**17**	

This type of visualization can be done for other factors such as time, case order, etc.
Differences in the complexity or difficulty of the cases can be reviewed at the individual item level; such differences, if intentional, are useful in designing realistic OSCEs that can differentiate among learners and also within an individual learner's strengths and areas for improvement

Internal consistency: Calculate measures of internal consistency for items designed to capture domains of performance (e.g., communication, history gathering, or differential diagnosis). While there are many ways to estimate reliability (see Revelle and Condon [11] for a comprehensive but practical review), based on the type of variable (ordinal vs. continuous), Cronbach's alpha provides a widely accepted, frequently used, and fairly robust method for describing item-construct correlation that is readily available in virtually all statistical software programs (but is not always appropriate, see McNeish [12] for a discussion of best practices for estimating internal consistency). Most often described colloquially as "how well items hang together," Cronbach's alpha can be viewed as a measure of the interrelatedness of items. The general idea is that if the items are highly interrelated, they can be thought to reflect the intended construct or area of competence.

Ordinal Cronbach's alpha and Cronbach's alpha ranges from 0 to 1, and a general minimum threshold of .80 is considered to represent minimal internal consistency. Ordinal Cronbach's alpha should be considered over Cronbach's alpha when there are fewer than five response options for an item.

If items were intended to be internally consistent and are not, this may be due to problems in the assessment such as poorly worded items, inaccurate application of items (e.g., differences between assessors or individual assessors not being consistent in how they interpret or apply the items to learners), external effects (e.g., disruptions during the OSCE), or contextual influences (e.g., case content or difficulty) on the assessment.

However, low internal consistency can also reflect problems with the theory underlying the assessment approach—that the domain the items were designed to assess is not, in reality, a stable, coherent, consistent construct as theorized. These questions of measurement theory are well beyond the scope of this chapter, and so we turn back to the practical. Note, just because an item does not hang together does not mean the item should be thrown out of the OSCE checklist but rather may suggest it is measuring some other unique contribution and thus should be evaluated in that context.

Most statistical programs will provide not only the overall Cronbach's alpha for all selected items but also, for each item, what the Cronbach's alpha would be if the item was NOT included. This can be very helpful for identifying problems with a specific item (whether with the item itself, how it was applied, or how well it actually fits with the other items)—which can then be more fully explored and improved upon in future iterations of the assessment. In the short term, those items that decrease the alpha to below the acceptable level can be dropped and not included in the calculation of domain summary scores.

2. Factor Analysis: Factor analysis can be used to determine the dimensions or factors within items [13] and is a useful complement to Cronbach's alpha in that it can help determine the dimensionality of constructs. If multiple factors are identified, the items are likely to be capturing multiple underlying dimensions of competence (or performance) that should be reported out separately. We describe factor analysis very broadly here and direct the reader to explore the appropriate factor analytic methods given the structure of the data (e.g., the use of weighted least squares approaches for ordinal data [14]. Factor analysis generally proceeds as either exploratory—where the goal is to identify possible factors—or confirmatory—where the goal is to confirm a priori "factors" based on conceptual models, previous data, or the literature. See Crutzen and Peters [13] and other available guides for full details. Figure 6.4 describes the underlying factor structure for our model of communication skills [15].
3. Computation of Scores: Once initial screening of items has been completed, the score itself can be computed. Scores can be calculated based on basic measures of central tendency, e.g., means when rating scales have meaningful numeric values or medians if data are not normally distributed or based on ordinal scales, or as total points (simple sum of "awarded" ratings), or as percent of possible "points" (how an individual was assessed compared to how they would have been assessed if they had been rated at the highest level for every item). We routinely use a mastery-oriented approach wherein we calculate the percent of behaviorally anchored rated as well done (on a scale of not done, partly done, and well done), in order to estimate a learner's ability to perform skills at the level needed to be maximally effective.

As noted in Chap. 2 ("Organizing OSCEs (and Other SP Exercises) in Ten Steps"), such scores can be calculated at the OSCE level (all items across all stations across all domains), the case level (all items across all domains within a station), or the domain level (a specific area of competence either across all cases or within a case). Each score reflects a specific inference.

OSCE-Level Scores: OSCE scores assume that the OSCE tells us something important about learners' performance aggregated to that level (a general state of competence that transcends the specifics of cases and domains) or where low performance in one area or case can be offset by high performance in another.

Case-Level Scores: Scores at the case level are meant to reflect the ability to handle all of the relevant dimensions of that scenario (such scores could help make decisions about whether learners could be entrusted to

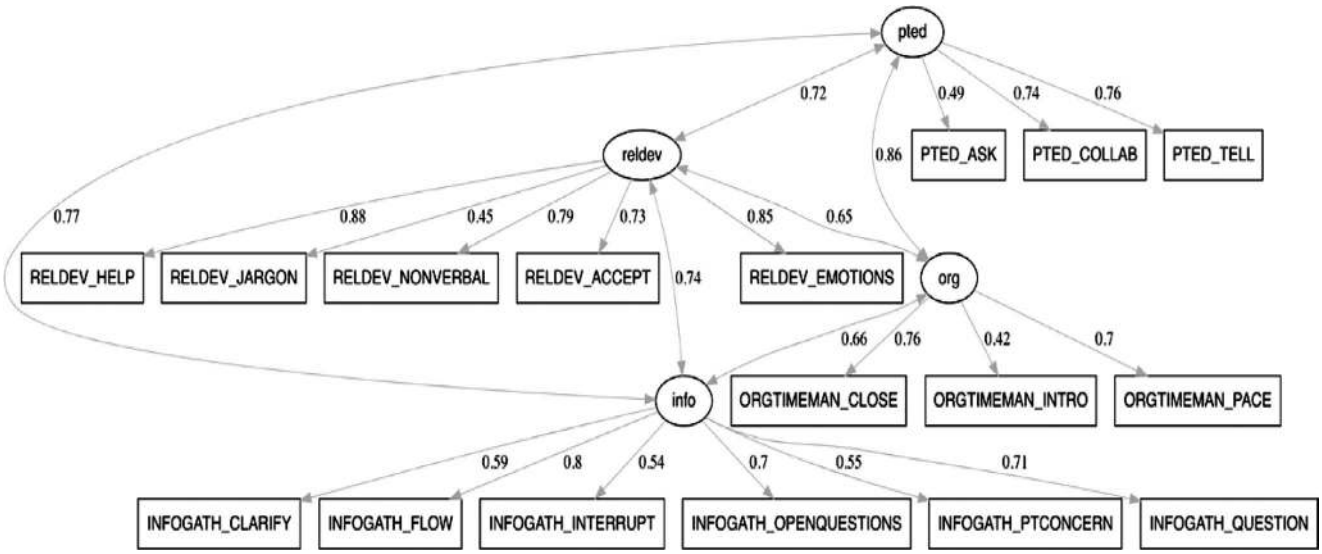


Fig. 6.4 Clinical Communication Skills Assessment Tool (CCSAT) confirmatory factor analysis. This figure [15] shows results of a four-factor confirmatory factor analysis for the Communication Clinical Communication Skills Assessment Tool (CCSAT). The factor loadings between items and subdomains, patient education (pted), relationship development (reldev), organization and time management (org), and information gathering (info) and correlations between the subdomains are shown

independently care for a real patient with the same clinical conditions and parameters in actual practice).

Category/Competence/Skill Area: Scores at the competence level are intended to represent competence in that skill area—usually with the underlying assumption that ability is best assessed across multiple samples of performance and therefore across multiple cases. However, an individual case can be used to establish competence for a very focused, defined set of related tasks.

As shown in Fig. 6.5, box plots are useful for describing many features of assessment data. Often known as box and whisker plots, they can depict the mean (X) and median (middle line of box) scores. More importantly, box plots visualize the variability in the sample data, including the interquartile range or the 25th (bottom line of box) to the 75th (top line of box) percentiles as well as pre-defined full variability (extent of the whiskers). The dots represent individual data points, and outliers are data points that fall outside of the whiskers.

Violin plots (Fig. 6.6), an enhanced version of a box plot, are particularly useful visualization techniques for exploring variation. Violin plots provide much of the same information that is available in a box and whisker plot—namely, the median, interquartile range, and upper and lower values—but provide more information on the distribution of the data, showing the distribution *shape* of the data. The width of the "violin" shape around the vertical line provides a visualization where the width reflects the frequency of data values at that point on the Y-axis, showing, for example, whether the distribution is bimodal or mostly clustered around a single set of values. Violin plots can also be used to show the change in distribution over time and can even be overlaid with individual learner change lines to show patterns of change over time.

4. Quality Assurance Analysis of Scores: Once scores have been computed (based on a priori conceptual frameworks, concept maps, the literature, and/or the results of the internal consistency and factor analyses described above), descriptive statistics can be used to illustrate and explore the score data. The goal of these analyses is to explore the distribution and variation of scores in the targeted samples (are the scores normally distributed or are they skewed high or low? Did most everyone perform similarly—as reflected in the score?). These types of analyses also begin to build evidence for (or against) the validity of the assessments and generally follow the approach of checking the operationalization of the assessment (what the collected and scored data looks like and how it relates to other data and factors) against what the theory would predict. For example, if scores were created based on the idea that they would reflect competency in a particular area for a group of well-prepared learners, then the validity of that assessment would be supported if the data was skewed to the right, meaning most learners performed well on the measure. But if you expected learners to have a wide range of abilities, you should be worried about your assessment if you found the same pattern of score results (skewed high). Below, we describe some specific ways to begin your exploration of the quality of your assessment data—note that each, when related back to theory-informed hypotheses about how the data "should" look, helps evaluate the validity of the assessment.

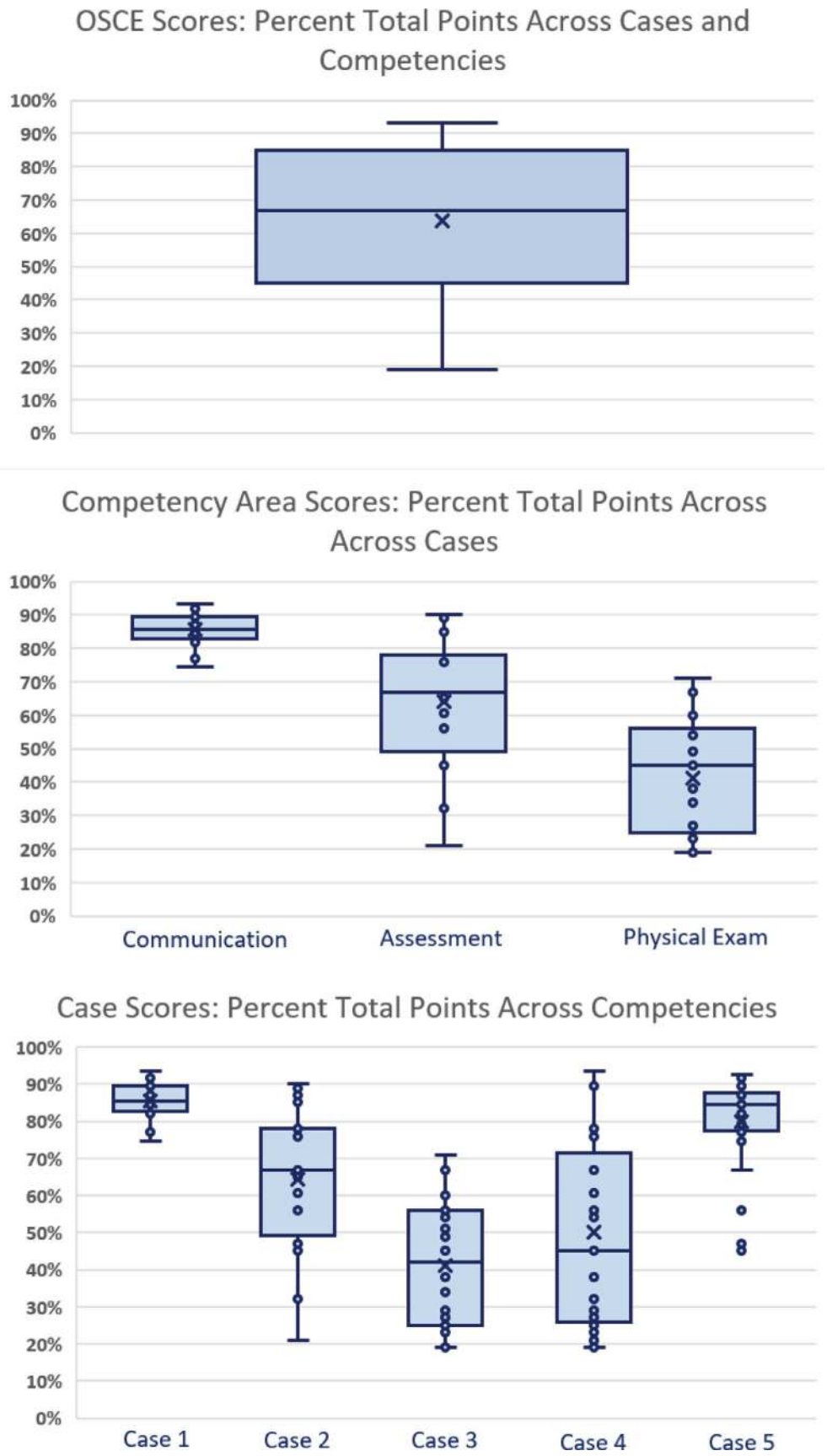


Fig. 6.5 OSCE-, case-, and competency-level scores (for a cohort of learners). Each view of the "scores" tells a different story. Viewing the *OSCE score* provides a snapshot of broad performance, but reviewing the *competency area scores* suggests substantial differences in categories of performance. *Case scores* show that overall performance varies by case

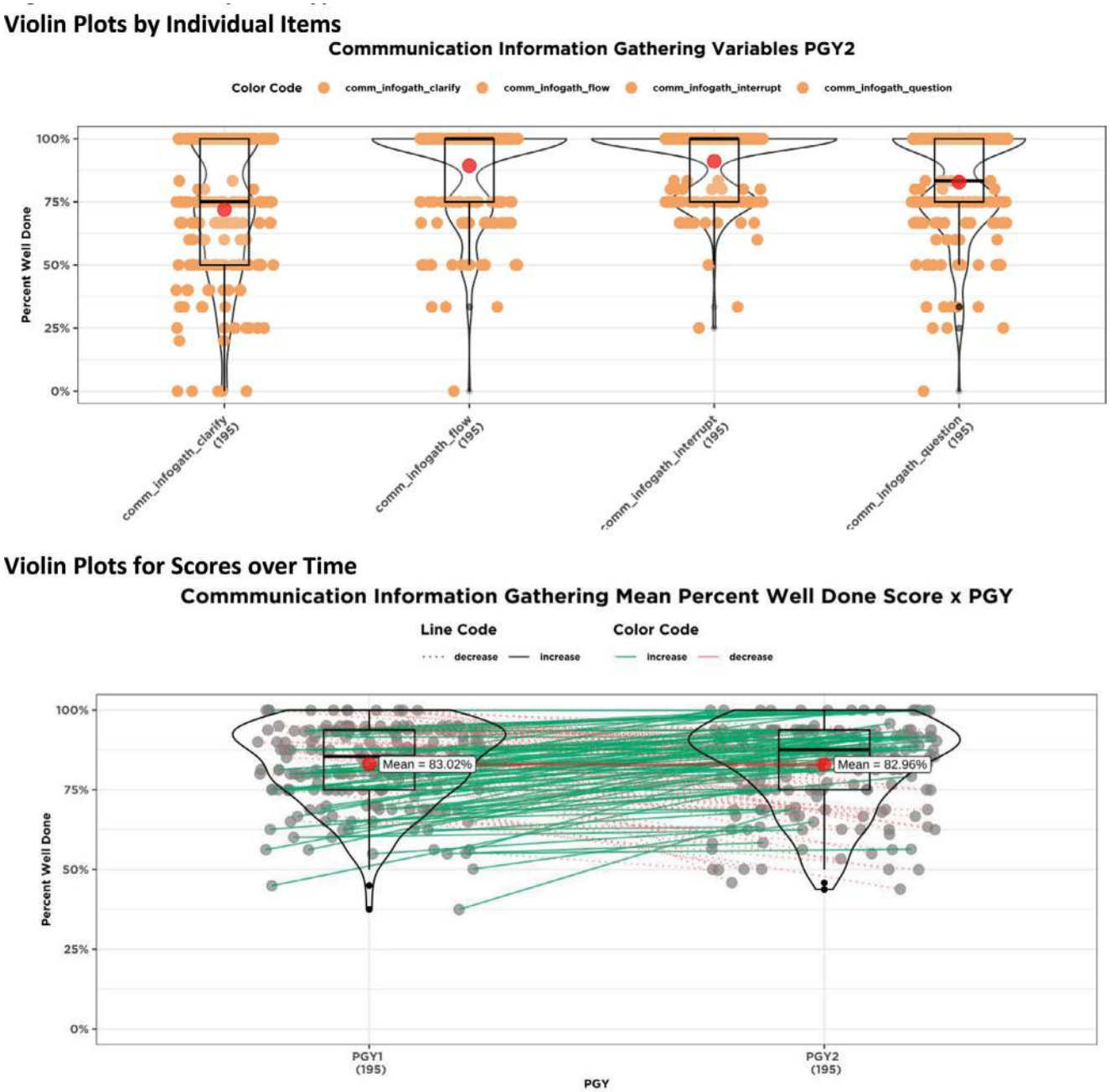


Fig. 6.6 Violin (density) plots for communication skills. The distribution shown in the violin plot for the "interrupt" item depicts a large number of "well done" responses (meaning that most learners did not interrupt), whereas responses were distributed more equally for the item "clarify" (meaning that learners were more equally distributed across the top half of the scale). The overall distribution of scores (shape of the violin) is similar for the two time points; however, by tracking the same learners from PGY1 to PGY2 and including individual change lines, we can see that some residents are improving while others are declining over time

Outliers: Screening for outliers—learners whose scores place them at the extreme edges of the distribution (very high or very low)—can suggest problems with the data (incorrect coding of values or calculation of scores or an assessment that is not accurate). Outliers can also help identify struggling or outstanding learners. Different software programs use different approaches to determine cut-offs for what constitutes a data point being an outlier, but they often use two standard deviations above or below the mean. Outliers should be reviewed carefully as they may not represent an anomaly, per se, but the ability of the assessment to discriminate among learners and identify very low or very high performers. As shown in Fig. 6.5, box plots are helpful for visualizing outliers

(although one should review how outliers are defined within the program being used to generate the box plot, to ensure accuracy).

Distributions (Frequencies): As noted above, visualizing the full range of scores via frequencies, histograms (Fig. 6.7), or other ways of examining the distribution is an important early step. Beyond determining if there are outliers (and if those outliers seem to reflect actual extreme performance), seeing the distribution of scores helps educators, learners, and program evaluators make sense of the data. But first, the distribution should be checked against what was expected—if it is different, then one needs to explore whether it is the hypothesis/theory (what was expected) that was off base or whether it is the assessment instrument (or some combination of the two!).

Rater Effects: Analyses of raters influences that were conducted at the item level can also be analyzed at the score level (Fig. 6.8). These analyses can help identify potential data quality problems that operate at the level of summaries of performance.

Differences in score means by raters (if learners are distributed fairly equally/randomly across raters) can identify patterns of assessment that suggest bias or rating tendencies that don't fully reflect actual performance. These "hawk" or "dove" tendencies (being a harsh rater or an easy rater) can be identified and then addressed through training or accounted for in the distribution of

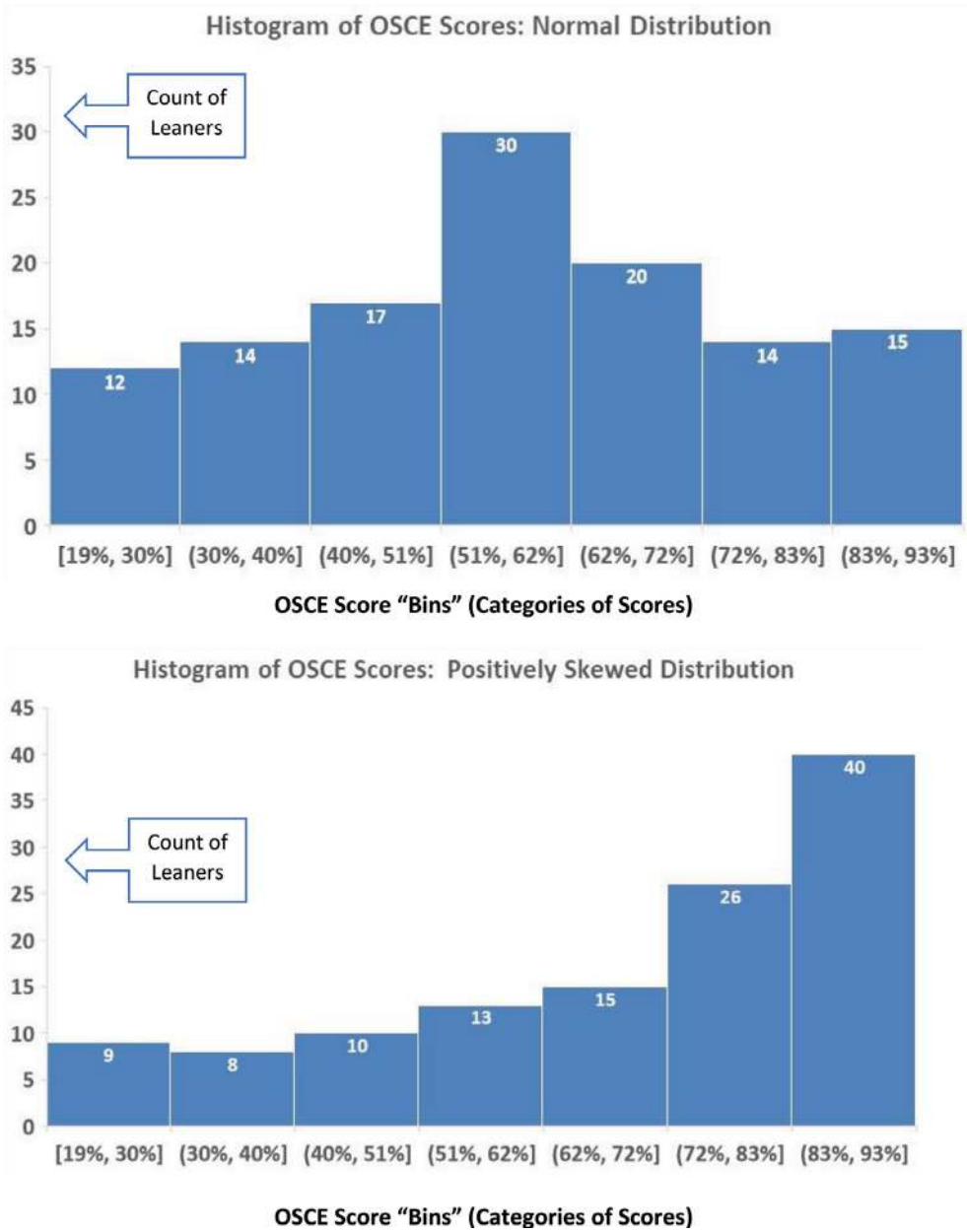


Fig. 6.7 Histogram of scores—showing different distributions. These histograms show counts of learners (Y-axis, height of bars) for score ranges (X-axis, each bar reflects a "bin" or range of scores). Note that a score of 54% in the top graph (normally distributed) should be interpreted very differently from the same score if it occurred in a sample in the bottom graph ("right-skewed" distribution). The bottom distribution might reflect an OSCE where most students are expected to achieve proficiency (set to be greater than 72%). While a 54% within the sample of learners described in the top histogram would reflect "average" performance, that same score among the learners described in the bottom histogram might suggest that student is not achieving expected levels of competence

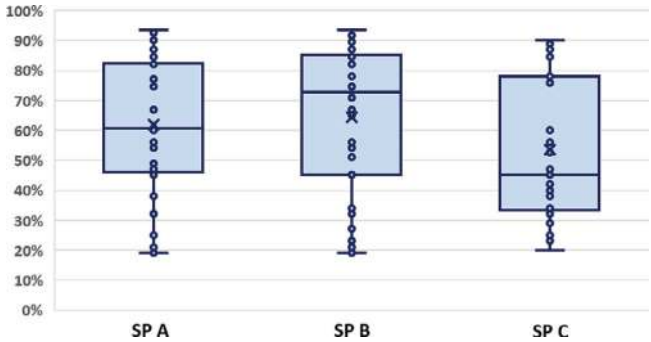


Fig. 6.8 Variation in communication scores by rater. SP A mean = 62.0% (SD 23.3%) n = 41; SP B mean = 64.2% (SD 25.3%) n = 43; SP C mean = 53.4% (SD 23.1%) n = 43. In this situation, SPs are each assessing one-third of a cohort of learners in the same case, assigned based on scheduling. The sample of learners for each SP is expected to have similar levels of performance. SPs A, B, and C show different distributions when their assessments are calculated as scores: note the variation in median and interquartile ranges. Comparison of the means using a one-way ANOVA finds that the SP means are not significantly different (F = 2.36, p = 0.09). Even though the means do not differ significantly, the results from the box plot suggest a closer review of the SPs' assessments to explore whether SP C is being a more stringent rater ("hawk") than SP A and B

raters. One can also look at the distribution and standard deviation of scores by raters, to identify raters with very little variation (suggesting over-consistency such that the ratings don't reflect the underlying variation) or more variation (suggesting that ratings are potentially arbitrary, again, not reflecting the underlying variation in the learners' performances) when compared to other raters. Examining rater patterns can also help determine if raters are using the full scale and also help facilitate future training with SPs if they are used repeatedly in an OSCE program, as quality assurance.

Case/Context Effects: Analyses of other contextual influences such as the day of the OSCE, time of the session, order of cases, and case-specific effects can also be explored at the score level. Case effects are an inherent challenge in OSCEs because of the challenge of case specificity: whether performance in a competency and/or skill area is likely to generalize (be consistent/similar) across clinical conditions and other aspects of the station (e.g., emergent vs. primary care) or if it is expected to be closely linked with the station characteristics (i.e., performance will depend on the contents of the case).

In terms of quality assurance, examining case effects helps determine the degree of case specificity found in an OSCE. The goals and design of the OSCE, of course, may determine how much case specificity is expected: were the cases designed to be similar in nature and therefore capture generalizable skills or were the cases meant to be distinct, capturing different skills sets requiring different knowledge bases?

In addition, however, analyzing scores at the level of cases can identify differences in the difficulty of cases—scores consistently higher on a case could suggest that the case was relatively "easy" compared to other cases OR that the assessment itself wasn't accurately capturing the complexity of performance (making it easy for everyone do well) OR problems with the raters being too "dove-like." Student variation can also provide insight into whether performance is case dependent or case-agnostic—showing case-specific ability versus more global ability. For instance, a student with a large variation across a series of cases that assess the same skills (such as communication) can suggest their ability to communicate varies by case-specificity. If students have very little variation this may suggest that performance is case-agnostic. Using variation and case performance can be essential in understanding our learners and help us design OSCEs with case selection in mind. Item Response Theory (IRT) function curves can be used to obtain case curves to see which cases are "easy" (differentiate learners with no ability from others) and which cases are "hard" (differentiate experts from all others). This can help design an OSCE with clear intention in terms of which levels of learners are differentiated.

Finally, quality assurance efforts can explore whether there are systematic differences in scores by other characteristics of the OSCE or contextual influences such as the time of day (morning vs. afternoon), order effects (e.g., case order or whether from a "test security" perspective, learners who participate later in the OSCE do better than the initial learners), or differences in setting (cases held in differently configured visit rooms or in simulation centers vs. other contexts).

Standard Setting: Making sense of scores requires educators to determine "standards" for performance: which level of scores are expected? Which are concerning? This is especially important when scores are used for summative judgements and grading decisions—judging an individual learner to be competent, entrustable, or ready to move to the next level. There is a rich literature and evidence base for setting standards that involve careful review of the assessment data (and often videotaped samples) to establish standards of performance through a combination of consensus and psychometrics. See Yousuf et al. [16], Danette and Norcini [17], Boursicot et al. [18], and Downing et al. [19] for more information.

Reporting and Visualization of Data

Overview

OSCE assessment data is only useful if it can be accessed and analyzed and the results visually represented and, as noted earlier, if users can have some trust and confidence in its accuracy. Earlier, we described the processes and pipelines that are needed to make sure OSCE data is of sufficient quality and readily available. Now, we shift to how to make sense of the data. Reporting is all about translating data in to insight and information. How data is visualized is key to ensuring everyone understands the information and can act on that shared understanding in a consistent way. In most academic institutions, a great deal of educational data is collected but not routinely given back to learners and educators in a timely, actionable, and interpretable manner. This means that learners can't adjust their learning and educators can't improve their teaching "just in time." The advent of sophisticated simulation software systems has helped many overcome this problem because such programs include the ability to provide learners and educators with views of completed assessment forms and with summary views of data. While completed assessment forms are helpful for the granular feedback on performance, they are not very helpful for understanding more broadly how learners' skills are developing holistically, across clinical conditions, between different domains of competence, and over time. Most simulation software programs have developed score reports to address these questions. However, due to the complexities of meeting a very diverse range of customer needs within one software program, reports are generally are somewhat limited in how scores are calculated and what kinds of analyses and visualizations are available.

OSCEs are flexible and, as is evidenced throughout this book, can be used in many different ways to achieve different educational, training, feedback, and assessment goals. Effective data analysis is also very complex—datasets are virtually always multidimensional, and while best practices have been identified, those are meant to be applied to very specific questions and data structures/types and adapted to the particular strengths and weaknesses of the data. One-size-fits-all data analysis and reporting just does not exist, especially in the context of OSCEs and the many ways in which they are used. At the same time, the number of tools available for analyzing and visualizing data has grown exponentially. While the business world has capitalized on those tools, the context of education is very different, and thus, tools for medical education still need to be developed on how best to harness our data with technologies for turning assessment data into knowledge and action. Too often, the OSCE assessment results we return to educators and learners are decontextualized from the point of assessment, oversimplified and reductionistic (telling only one part of the story or performance, underrepresenting the true complexity of skills), presented in isolation, and delivered too late for impact. In general, one can start with a simple, fairly high-level summary view and then can drill down to greater granularity, including performance in different contexts (domains, cases) and over time. Thus, visualization of OSCE data requires multiple ways of looking at the various levels in order to be useful.

Reporting of data can take many forms, including, but not limited to, data visualizations—graphical representations of information using visual elements like graphs, charts, pictures, maps, etc. Visualizing data is both a science (with evidence-based principles) and an art, and this chapter will not even begin to tackle the breadth and depth of best practices in data visualization. Instead, we seek to illustrate some of those principles by providing examples of visualizations linked to specific reporting approaches and educational and training goals. By providing this menu of visualizations, we hope to stimulate ideas for reporting.

Moving forward, medical educators can take advantage of the broad range of tools now available to design and implement visualization and reporting systems that are educationally meaningful, that support skill and competence development, and that can be used to ensure that learners and trainees are prepared to be the best healthcare providers that they can be. Key to achieving this is following the principle of backwards design. This means starting the design process with the end in mind: providing students and educators easily understandable data that allows them to recognize strengths, make learning plans, and design and/or implement curriculum. We need to ask ourselves: what data is educationally meaningful and how to best display it? What is the goal of the analysis and reporting? How do we intend for the data to be used? For whom is the information directed? What should be communicated about the nature of the assessment data itself? In all this, it is important to engage our stakeholders early on, test our visuals, and develop a narrative and assurance is another area to consider, but will not be covered in this chapter.

Goals and Types of Reporting

A relatively simple way of thinking about reports for OSCEs is to consider for whom the report is designed. Is it for the learners, the educators who designed and implemented the OSCE, or for those responsible for evaluating the curriculum within which OSCEs reside? However, the goal of the OSCE assessment should also be considered: is the OSCE designed to provide formative feedback for learning, part of the cycle of practice, assessment, and feedback core needed for skills acquisition and mastery? Or is the OSCE designed to permit

summative judgements of the competencies of individual learners based on their performance of defined tasks in specific contexts? Is the OSCE meant to provide cohort-level information on the effectiveness of a curriculum or training program? Reports can also address the level of the "construct" being assessed—is the OSCE meant to capture specific practices (e.g., adherence to guidelines such as hand-washing or evidence-based medicine), behaviors associated with developing competence, demonstration of the skills needed to make judgements that learners can be entrusted to perform the patient care task in actual practice with limited supervision, or mastery of a complex set of skills needed to move to the next training stage?

Just to increase the degree of difficulty, many OSCEs are, and arguably should, be used to address the needs of multiple "users" and answer multiple questions simultaneously! A single report or reporting strategy cannot meet all those needs. Each combination of user, question, and level of inference about the underlying phenomenon being assessed requires careful design of a targeted reporting approach. Once the approach is fully specified, then decisions about how best to visualize the data can be made.

Reporting for Learners

Reports provided directly to learners often focus on helping learners understand "how they did" (i.e., did they do well?). What counts as "good" is often defined by normative comparison with a defined group of similar learners (e.g., class, cohort, training program).

Normative comparisons are based on locating an individual learner's performance within the performance of a group of peers—did the learner perform better, the same, or worse than others? See Fig. 6.9 for an example. The mean is often the basis for comparison with the focus on whether the learner's scores were above the mean, around the mean, or below the mean. Many reporting programs use standard deviation to determine when an individual learner's score places them substantially above or below the mean. Thus, a report might suggest that a learner who performed more than two standard deviations above the mean is "exemplary" while one whose score places them two standard deviations below the mean is likely in need of remediation. This is based on the characteristics of a normal distribution in which most (two-thirds) of the responses in this "normal curve" fall within ± two standard deviations of the mean. This can be a helpful way to display relative performance. Perhaps too often, however, we use these kinds of normative comparisons because they are easy to compute and require no additional decision-making or review.

However, whether normative comparisons are helpful depends on (a) whether the assessment data is normally distributed (either in actuality or is expected to be normally distributed) and (b) the goal of the assessment. If the primary goal is to provide feedback to the learner as to whether they are "on track" or not, then normative comparisons are only helpful if we expect the comparison cohort to represent a normal curve (bell-shaped, most learners in the middle clustered around the mean, and some at the bottom and some at the top of the distribution) AND if "on track" is well represented on that curve. But if all learners are doing well, then a normative comparison may set unrealistic expectations. A learner could be "on track" and doing fine but if compared to a skewed sample could receive the (inaccurate) message that they are underperforming. Given the selection biases built into medical school and other health professions education programs, we may often be analyzing performance data that is highly positively skewed (many individuals performing well). Learners also have expressed concern that comparing

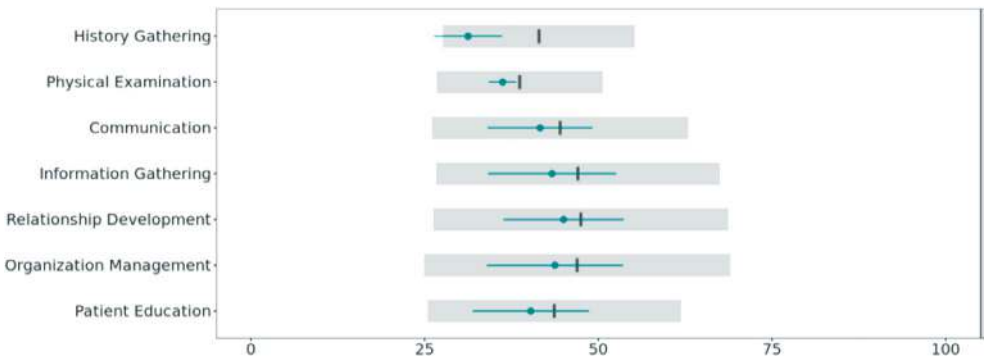


Fig. 6.9 Example of normative comparisons. Providing learners with information about how they performed in comparison to their class (or cohort or peers) can help provide context

The gray vertical line is the mean for your class, and the gray bar represents the class range (1 standard deviation above and below the class mean)

Your mean is represented by the turquoise dot, and the turquoise line represents your range (1 standard deviation above and below your mean)

The chart above shows a bar chart for each learner in the cohort, sorted from low to high, and locates the individual learner (indicated by the gray bar and star) within that range

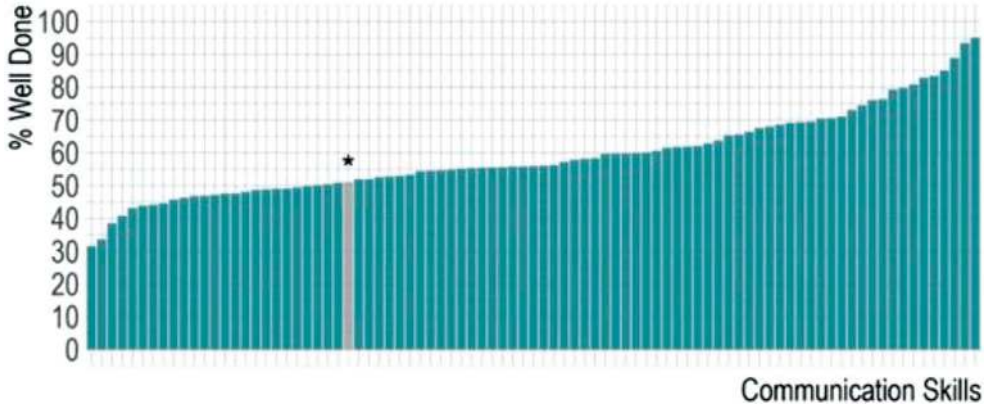


Fig. 6.9 (continued)

performance amongst one's peer implicitly (and sometimes explicitly) encourages competition and undermines collaborative learning and teamwork.

Formative Assessment for Learners

Formative assessment is generally viewed as assessment *for* learning—a way of providing targeted feedback to learners (and their educators, coaches, and advisors) that directly suggests areas of strength and areas of improvement and that is sufficiently granular/specific in order to provide actionable information on *how* to improve in weak areas (and sustain strengths). If we return to our consideration of the level of assessment (items, cases, categories of skills, OSCEs, etc.), reports that show learners how they were rated or assessed on **each individual item** are often used for formative assessment (Fig. 6.10. Items are usually very specific (focused on discrete, well-defined practices or skills), and if available rating options reflect behaviorally grounded observable examples, item-level analyses can provide clear information on how a learner could have performed more effectively.

Reviewing performance on specific items *across* different cases or clinical scenarios (Fig. 6.11) helps learners diagnose whether they are struggling with particular skills no matter the context or if it's only in certain situations (complex medical conditions, difficult or emotional encounters) that they struggle. While normative comparison has its limitations (described above), here, it can be very helpful to understand whether most learners at the same training level demonstrated similar patterns of performance. If so (especially if there is evidence from similar cohorts over time), then the learner is likely to interpret the results as suggesting that they are performing as expected and while this might be an area for improvement or goal-setting, it is not of concern. If the learner sees that most learners are able to demonstrate a skill that they cannot, then they may naturally be concerned—however, that is where an advisor, coach, or educator can step in to help the learner understand and analyze their performance.

In addition, however, having access to views of the assessment data that provide more of a complete picture is also essential. A sample view of the NYU Grossman School of Medicine OSCE dashboard is shown in Fig. 6.12. These views can be dynamic, so the learner (and/or their coach) can select that item and view their performance across cases (e.g., this is an item that only causes me problems when I don't have a solid foundation of medical knowledge), over time (e.g., I see that most learners acquire this skill by the end of the year and I'm already noticing improvement from my previous OSCE), within other learner cohorts (e.g., is this something learners at the next stage master?), or in association with other items (e.g., do learners who don't perform well on this item also struggle with specific other items and, if so, what category of skill does that represent?).

A static report can also depict this information to help the learner (and their coaches) analyze the situation. See Fig. 6.13 for some excerpts from an individualized report that is created for each student through the use of R and the Shiny App package. It includes both high-level summative aggregate performance information (the four core clinical skill scores used to make pass/fail decisions in the top visualization) but also much more granular data on performance that includes more context and provides more actionable feedback (the drill down of the specific history gathering skills by case and specific history domains shown in the bottom visualization).

Whether to design dynamic, dashboard-based approaches or static reports depends to some degree on the data literacy of the users. Working within a dynamic dashboard often requires some basic comfort and familiarity with both technology and data. It also requires that learners be engaged in and curious about their assessment data. Unmotivated learners are unlikely to take the time and invest the cognitive effort to review and analyze their data

COMMUNICATION SKILLS

INFORMATION GATHERING

1. **Introduction**
   - ○ NOT DONE: Did not introduce self
   - ○ PARTLY DONE: Did not provide a complete introduction
   - ● WELL DONE: Provided a complete introduction that included FULL name, training status, purpose of interview AND referred to SP with full name or surname
2. **Comment 1 (ONLY CHECKED COMMENTS APPLY)**
   - ☐ First name only felt unprofessional
   - ☐ I didnt know your status or where you fit into the team
   - ☐ I felt that I was given no choice about talking to you
   - ☐ When you stated your name, your back was to me and you were shutting the door, so it was hard to hear
   - ■ None selected
3. **Use of patient-centered, open-ended questioning**
   - ○ NOT DONE: Started with closed yes/no questions
   - ○ PARTLY DONE: Began with an open-ended question but conducted most of the encounter with closed-ended questioning
   - ● WELL DONE: Started with open-ended questions and continued using them as appropriate throughout the interview
4. **Comment 3 (ONLY CHECKED COMMENTS APPLY)**
   - ■ Open ended questions made it easy to tell my story
   - ☐ You started with an open-ended question but the rest were closed
   - ☐ Closed-ended questions made it difficult to tell my story
   - ☐ I didnt feel like you were hearing what I had to say
   - ☐ When you started talking about yourself, the focus switched to you
   - ☐ None selected
5. **Elicitation of patient's thoughts on symptoms and underlying concerns**
   - ○ NOT DONE: Did not ask
   - ○ PARTLY DONE: Asked but did not provide appropriate response to encourage full disclosure
   - ● WELL DONE: Asked so that could fully share thoughts and concerns and provided response that encouraged full disclosure

Fig. 6.10 Formative assessment—completed checklist

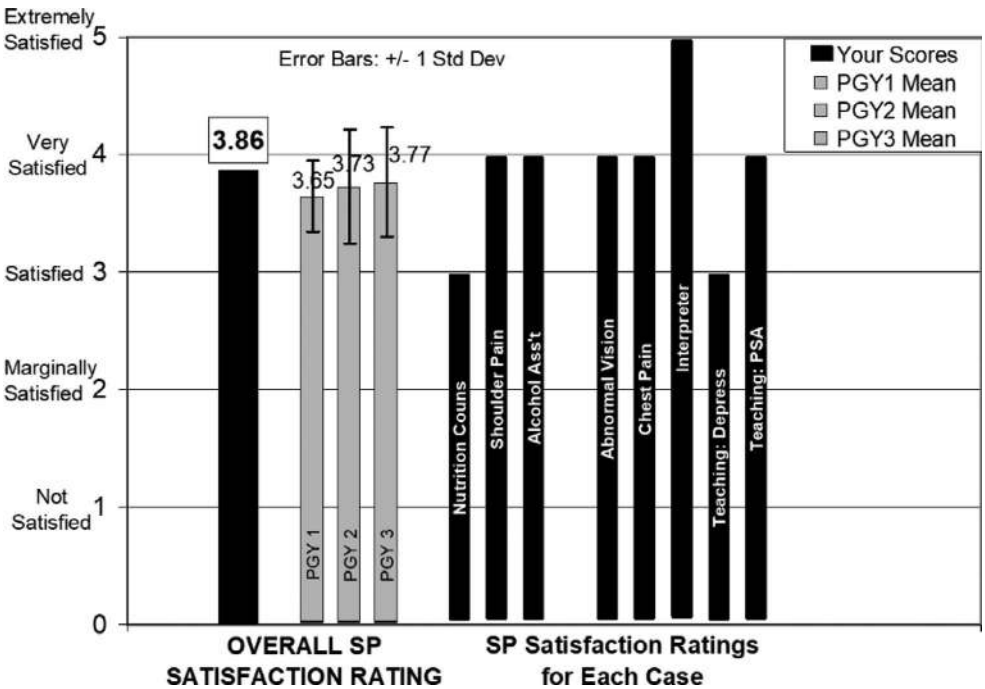


Fig. 6.11 Formative assessment of performance across cases and with comparisons. Notes: This formative assessment report is provided to individual residents so they can see their overall performance in a category (in this case, patient satisfaction), compare it to their peers by postgraduate year level, and then see how it varied across the cases in which it was assessed. This resident generally received high satisfaction ratings; however, ratings were lower in two cases, and the resident could reflect on what may have led to decreased patient satisfaction in those contexts

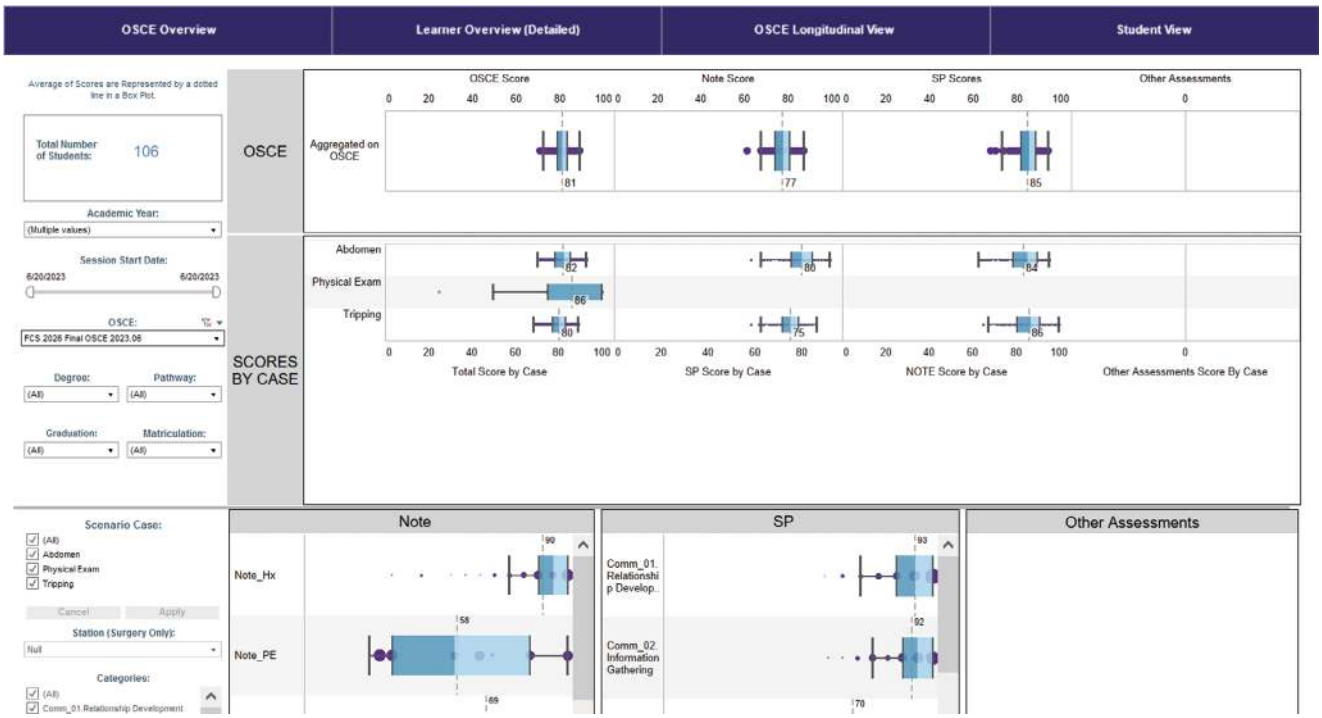
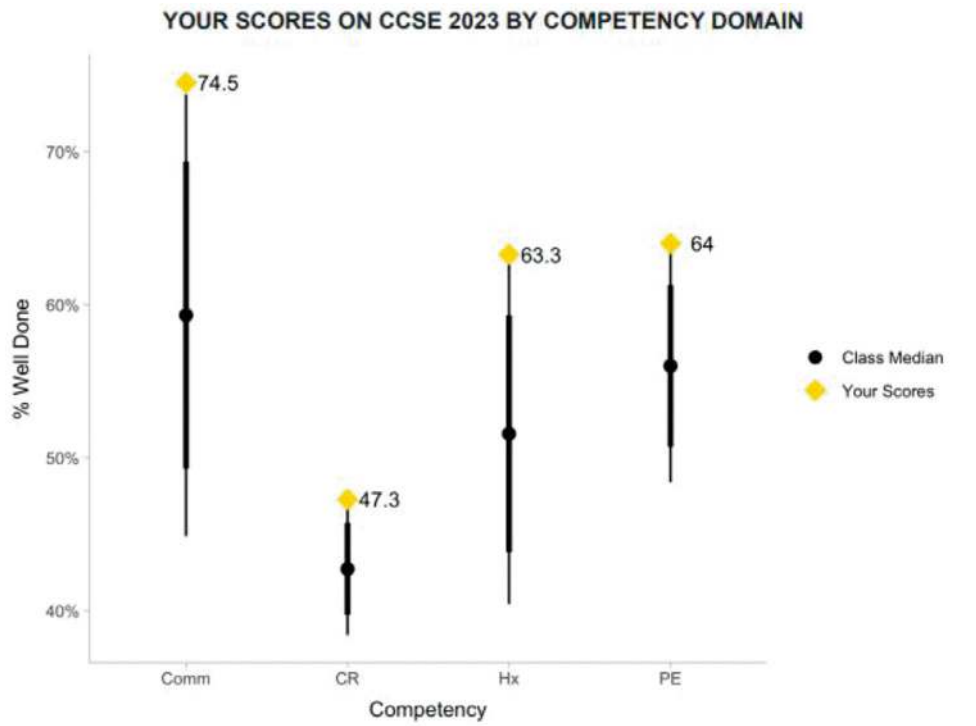


Fig. 6.12 NYU Grossman School of Medicine OSCE dashboard. This dynamic dashboard is built using Tableau and based on a data feed from our simulation center. The view featured above provides an overview of all OSCE data with filters for year, OSCE, cases, assessment types (Note, SP, Other), and categories (competencies, EPAs, etc.). Additional views provide a Learner Overview, Longitudinal View, and Student View. See Figs. 6.17 (longitudinal learner view) and 6.24 (curricular evaluation and planning view) for examples of additional views provided by this OSCE dashboard

interactively. Educators can enhance motivation by combining review of the data with reflections, advisory/coaching meetings, or preparation for upcoming high-stakes assessments. They can promote engagement and curiosity through small group sessions or by posing questions about the data that learners can seek to answer. Making assessment data more transparent and providing learners with greater control over their assessment data can also lead to more engagement and more thoughtful analysis and reflection. It's important to recognize that testing visualizations with learners is crucial to ensure they accurately interpret the graphs and grasp the intended message. Overloading dashboards with too many interactive elements can lead to misunderstandings, while imposing too many constraints may discourage use if learners feel they can't interact and uncover additional insights beyond a static image.

On the other hand, reports can be designed to be more selective and focus the learners' attention on specific ways of viewing the data (and therefore, specific interpretation guidelines as to what matters). It is also usually easier to include explanations, interpretations, and additional information in reports than it is in dashboards due to the challenges of limited space available on our digital screens, especially when viewed on smart phones or tablets.

While the rich literature on feedback cannot be accurately represented in this chapter, the core principals of providing feedback are that it should be "timely, specific, actionable, and task-oriented rather than person-oriented" and should help learners understand how to work toward improving their skills [20–22]. While we are currently focusing on the content of the feedback, it is nonetheless essential to take the context and the relationship between the individuals receiving and giving feedback into account (see for example the R2C2 feedback model ([23, 24]; and [25]) and theories of education alliance [26]). Designing reports and dashboards should ideally take this context very seriously: will learners be viewing the dashboard on their own or with a trusted advisor? How well do learners understand the results? Can learners use the results to develop action plans to improve weaknesses and sustain strengths? Will the results motivate learners, make them complacent, or promote a sense of helplessness? All of these questions (and more) can be explored throughout the development and design of these data visualizations—inviting stakeholders to reflect on and respond to the various ways of representing OSCE data, exploring the specific context of the data and its place in the curriculum/training program, and investigating the effects of feedback on the learners. This can be done through design thinking plan-



ADDITIONAL FEEDBACK BY EPA DOMAIN

Global Assessment Scores by EPAs

In each of the cases, Standardized Patients provide a global Entrustable Professional Activity (EPA) assessment of communication/patient education skills (EPA11) and note raters provide a global assessment of history (EPA 1a), physical exam for 4 cases with an exam (EPA1b), clinical reasoning by domain (EPA2, EPA3a, EPA3b), and documentation skills (EPA5). Below you will find the summary of those global assessments across the CCSE.

EPA	Poor	Beginning	Competent	Strong	Total
EPA 11 - Communication (SP)	0	0	1	5	6
EPA 1a - History Documentation (Note)	0	3	2	1	6
EPA 1b - Physical Exam (Note)	1	1	1	1	4
EPA 2 - Prioritized Differential Diagnosis (Note)	0	2	1	2	5
EPA 3a - Diagnostic Test Selection & Interpretation (Note)	0	2	2	2	6
EPA 3b - Management (Note)	0	2	2	1	5
EPA 5 - Documentation (Note)	1	1	2	2	6

Fig. 6.13 NYU Grossman School of Medicine CCSE report. This individualized report is created for each student through the use of the statistical programming language R and the Shiny App package. It includes both high-level summative aggregate performance information (the four core clinical skill "scores" used to make pass/fail decisions in the top visualization) but also much more granular data on performance that includes more context and provides more actionable feedback (the drill down of the specific history gathering skills by case and specific history domains shown in the bottom visualization)

HISTORY-EPA 1a (Assessed by Standardized Patients and Trained Note Raters)

% Well Done for history gathering (assessed by standardized patients) and history documentation (assessed by trained note raters) for each case, as well as across all 6 cases

	Abdominal Pain	Cough	Falls	Reproductive Health	Palpitations	Pediatric	OVERALL
History Gathered (SP)	100.0%	42.9%	100.0%	100.0%	100.0%	100.0%	**90.5%**
History Documentation (Notes)	45.5%	50.0%	50.0%	46.2%	41.7%	41.7%	**45.8%**

% Well Done for each checklist item across cases

Assessment Input	ITEM	% Well Done
Notes	Chief Concern	50.0%
Notes	Baseline Health	60.0%
Notes	Attributes of the Symptom	16.7%
SP	Attributes of the Symptom	77.3%
Notes	Risk Factors	40.0%
SP	Risk Factors	92.0%
Notes	Pertinent Negatives	66.7%
Notes	Time Course	40.0%
SP	Time Course	100.0%
Notes	PMH	44.4%
Notes	PSH	50.0%
Notes	Social History	33.3%
Notes	Family History	50.0%
Notes	Medications	66.7%
Notes	Allergies	33.3%
SP	Medications/Allergies	80.0%

Fig. 6.13 (continued)

ning sessions, focus groups or informal interviews on initial responses and reactions, or ongoing post-OSCE surveys or questionnaires that elicit feedback on the data dashboard or reports. These Continuous Quality Improvement (CQI) processes are essential to effective reporting—not only should reports be designed based on the comprehensive literature on feedback and the context but ongoing monitoring efforts must be in place to ensure the reports are achieving their intended goals.

Of course, not all dashboards are dynamic—they can be designed to be fairly static, with defined views of the data, and limited opportunities for querying or filtering the data. And some reports (for example, html-based Shiny App reports) can include limited interactivity. Regardless of format, it is important to consider learners' motivation, engagement, focus, and data literacy in deciding between interactivity and fixed reports.

Applications that allow users to visualize and manipulate the data as part of the process create reports that reflect the underlying OSCE data and incorporate expert judgement about the meaning and interpretation of the data. Figure 6.14 is a screenshot from our Comprehensive Clinical Skills Exam "CCSE Grader" which allows faculty experts to review the OSCE data for a cohort of learners and determine appropriate pass/fail thresholds for a high-stakes multiple-station OSCE in medical school. While the CCSE OSCE is standardized

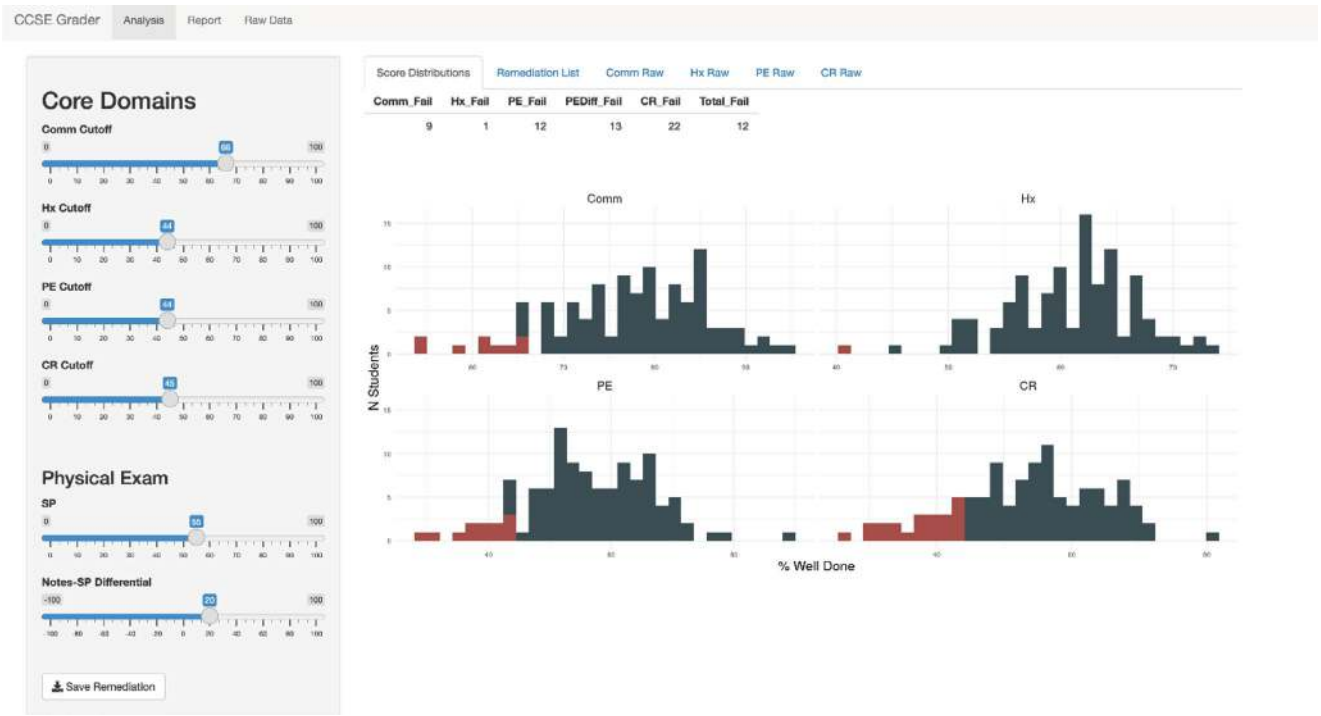


Fig. 6.14 Shiny App OSCE interface for determining pass/fail thresholds. The "CCSE Grader" is a Shiny App (R) interface that allows faculty experts to review the OSCE data for a cohort of learners and decide on appropriate pass/fail thresholds. Once the thresholds are determined, reports for each learner can be created and shared. This approach allows for near real-time analysis and review of OSCE data, to ensure the quality of the data and support timely decision-making and dissemination of accurate and meaningful results

and each administration is designed to be as similar as possible to previous administrations, it is also updated and redesigned as needed to align with student levels of competence and changing expectations. Therefore, decisions about how learners performed are best made with the actual data in hand. However, given the need to rapidly make decisions about remediation, training programs rarely have the luxury to conduct in-depth analyses, and so we use R and Shiny to create a dynamic set of views so that faculty experts can review the OSCE data for a cohort of learners and decide on the appropriate pass/fail thresholds. Once the thresholds have been determined, reports for each individual learner can then be created and shared. This approach allows for near real-time analysis and review of OSCE data, to ensure the quality of the data and support timely decision-making and dissemination of accurate and meaningful results.

Summative Assessment for Learners

Summative assessment usually occurs at the end of a training program or curricular unit and is designed to determine whether the learner has achieved specific learning goals or objectives. At the cohort level, summative assessment results can help evaluate the effectiveness of the educational or training program. The formative/summative dimension often also follows along the ungraded/graded distinction, respectively. Formative assessments can be viewed as part of the learning process—the often cyclical, iterative process of learning something new, practicing it, receiving feedback, refining that new skill, receiving additional feedback, and ultimately, perhaps, mastering the skill. Since the assessments are essential to learning in this context, how a learner performs on the assessment is usually not used to make judgements about their competence or level of performance—these assessments do not "count" and are not usually incorporated into grading or summative assessments. What matters most in formative assessments is how learners respond to these assessments: the degree to which they engage in a growth mindset or master adaptive learning wherein they use assessment data to set goals, work toward achieving those goals, and then use assessments to determine if they've met the targeted goals.

Summative assessments, on the other hand, tend to be more fixed and limited—they shed light on where a learners' skills are at a given point in time. While formative assessments often focus on more granular levels (items/defined skills), the level of analysis for summative assessments is usually more abstract or global (multiple items intended to capture a domain or construct), given that summative assessments are used to make determinations about a learner's mastery of skills or degree of competence. In addition, they require more samples of performance and more rigorous methods in order for educators to have confidence in the sum-

STUDENT NAME

Overall Performance: PASS

Reporting Guide

Communication

Standardized patients in each of the eight scenarios rated student communication skills, based on four components of the medical interview: organization, information gathering, relationship building, and patient education and counseling. Scores are reported as percent items from a standardized communication checklist rated ***done well***.

History Gathering

Standardized patients rated student history gathering skills using content-specific history checklists developed by the relevant clerkship director(s). Scores are reported as percent of items from the history checklist rated ***done well***.

Physical Exam

Standardized patients recorded whether each student performed the expected physical exam maneuvers and the level of skill with which the exam was performed based on specific behaviors defined by the clerkship directors. Each standardized patient was trained in the physical exam by a clerkship director or an experienced NYU attending designated by the clerkship director. Overall scores are reported as percent of expected physical exam maneuvers ***done well*** by the student.

Physical examination scores are also provided for each case as such skills may be quite specific.

Clinical Reasoning

You were asked to write patient notes after each of 5 cases. Faculty, using a standardized grading rubric, rated the clinical reasoning demonstrated in each of those notes, focusing on organization, identification of relevant findings and differential diagnoses. Faculty then rated overall clinical reasoning on a 4-point scale (1=poor; 2=fair; 3=good; 4=excellent). These overall ratings were averaged across the 5 cases to arrive at a summary Clinical Reasoning score.

We also provided you with information on the % of accurate findings noted, the total number of reasonable differential diagnoses, and the number of notes rated by faculty as "organized."

How to Interpret the Results:

Your individual scores in each of the above skill areas are indicated in the following graphs. To help you interpret this information, we have provided you with your score* as well as the mean score of your classmates with an error bar indicating the range of scores one standard deviation ***above*** the mean and one standard deviation ***below*** the mean.

Fig. 6.15 Example of summative assessment report for learners

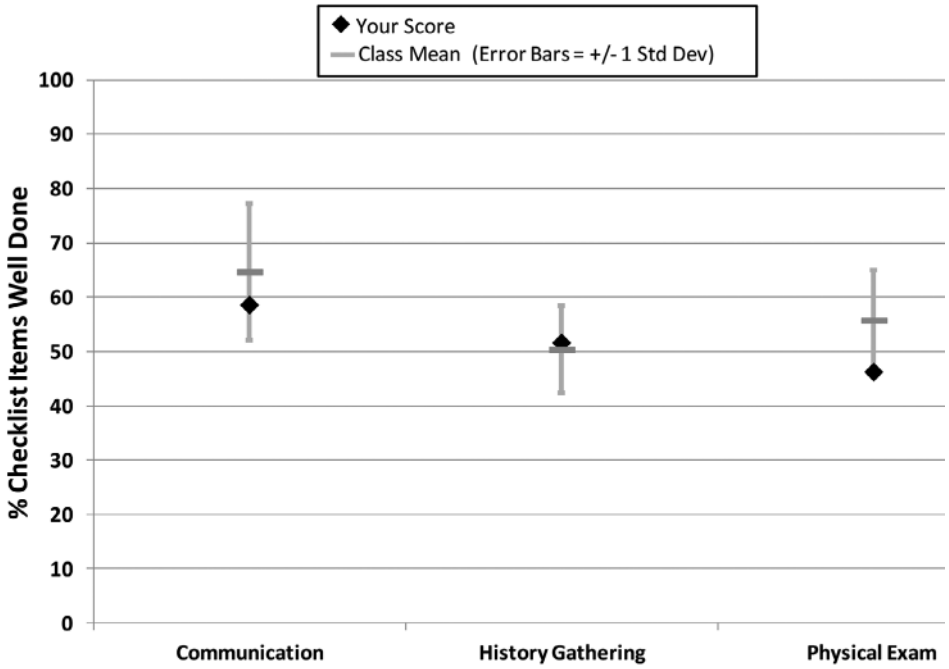
Based on careful analysis, student performance may be interpreted as follows:

- Students performing **one or more standard deviations above the mean** likely possess excellent skills and will likely perform well on similar tests in the future.
- Students **within one standard deviation of the mean** likely possess at least adequate clinical skills and are likely to pass similar tests in the future, but they may have specific areas in need of improvement.
- Students **with scores more than one standard deviations below the mean** possibly have weak clinical skills, as measured by this type of exam, and will benefit from an increased effort to improve such skills through practice, or, in some cases, specific remediation. Anxiety may account for poor performance on the exam. Practice may help this situation.

* % Well done = % of rated items for which you received a well done (vs. not or only partly done). A 0% means that you didn't get full credit for any of the assessed items. Missing data is indicated by **N/A** for Not Available.

YOUR SCORES ON THE EXAM

SAMPLE STUDENT: Communication, History Gathering and Physical Exam Scores



Due to the specificity of physical examination skills, we also provide you with your % well done score for the assessed physical examination skills within each of the 5 physical exam cases:

Physical Examination Case Specific Scores	YOUR SCORE % Done Well	CLASS MEAN % Done Well	CLASS RANGE 1 Std Dev Below Mean	CLASS RANGE 1 Std Dev Above Mean
Acute Abdominal Pain	**56%**	69%	54%	84%
Shoulder Pain	**72%**	53%	34%	72%
Palpitations	**45%**	46%	25%	67%
Vision Loss	**59%**	54%	41%	67%
Dyspnea	**48%**	53%	37%	69%

Fig. 6.15 (continued)

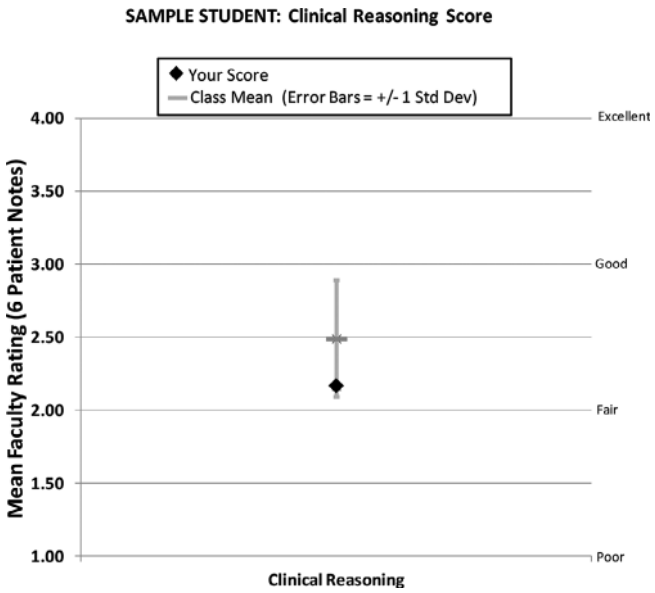
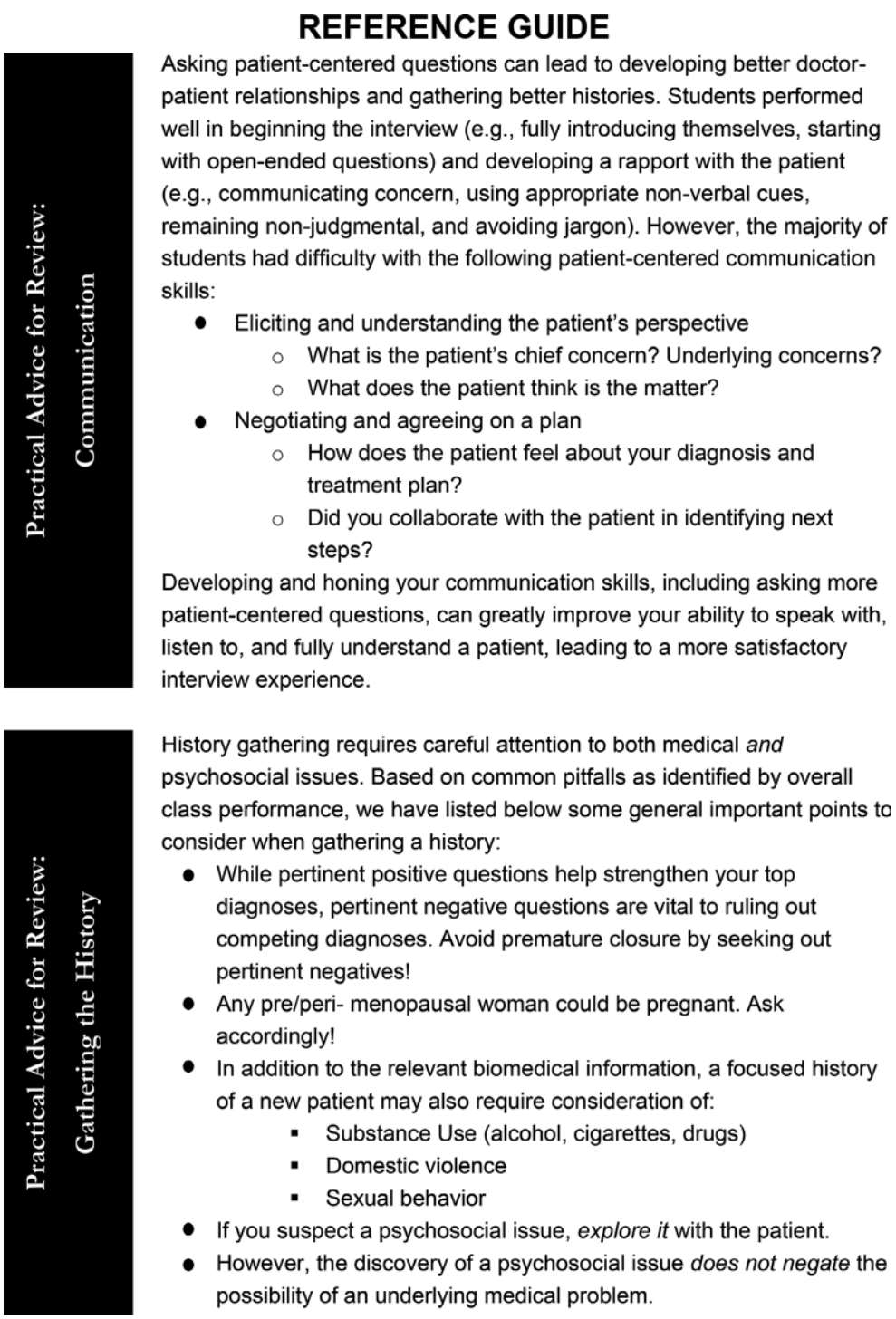


Fig. 6.15 (continued)



REFERENCE GUIDE

Practical Advice for Review: Communication

Asking patient-centered questions can lead to developing better doctor-patient relationships and gathering better histories. Students performed well in beginning the interview (e.g., fully introducing themselves, starting with open-ended questions) and developing a rapport with the patient (e.g., communicating concern, using appropriate non-verbal cues, remaining non-judgmental, and avoiding jargon). However, the majority of students had difficulty with the following patient-centered communication skills:

- Eliciting and understanding the patient's perspective
  - What is the patient's chief concern? Underlying concerns?
  - What does the patient think is the matter?
- Negotiating and agreeing on a plan
  - How does the patient feel about your diagnosis and treatment plan?
  - Did you collaborate with the patient in identifying next steps?

Developing and honing your communication skills, including asking more patient-centered questions, can greatly improve your ability to speak with, listen to, and fully understand a patient, leading to a more satisfactory interview experience.

Practical Advice for Review: Gathering the History

History gathering requires careful attention to both medical *and* psychosocial issues. Based on common pitfalls as identified by overall class performance, we have listed below some general important points to consider when gathering a history:

- While pertinent positive questions help strengthen your top diagnoses, pertinent negative questions are vital to ruling out competing diagnoses. Avoid premature closure by seeking out pertinent negatives!
- Any pre/peri- menopausal woman could be pregnant. Ask accordingly!
- In addition to the relevant biomedical information, a focused history of a new patient may also require consideration of:
  - Substance Use (alcohol, cigarettes, drugs)
  - Domestic violence
  - Sexual behavior
- If you suspect a psychosocial issue, *explore it* with the patient.
- However, the discovery of a psychosocial issue *does not negate* the possibility of an underlying medical problem.

Fig. 6.15 (continued)

Practical Advice for Review: The Physical Exam

A thorough and systematic physical examination will provide the highest yield in obtaining accurate and complete clinical data. Based on frequent omissions and inappropriate techniques identified by overall class results on the physical examination, we have accumulated some advice for improving your physical exam skills:

- Remember, a thorough cardiopulmonary and abdominal exam includes inspection, auscultation, palpation and percussion!
- The fundoscopic exam requires practice until you can consistently see the retina. A good reference is:
  Bradford CA. *Basic Ophthalmology*. 7th edition. San Francisco, CA: American Academy of Ophthalmology, 1999.
- A comprehensive cardiovascular exam specifically includes auscultation for bruits as well as examination of the extremities for peripheral pulses, edema and color.
- Review non-routine maneuvers used to assess for specific diagnoses such as psoas sign and obturator sign. Know when and how to perform such maneuvers.
- Take advantage of the 4th year of medical school and practice, practice, practice! During this time, carefully review your comprehensive physical exam skills including abdominal, cardiovascular, musculoskeletal and neurological exams.

Fig. 6.15 (continued)

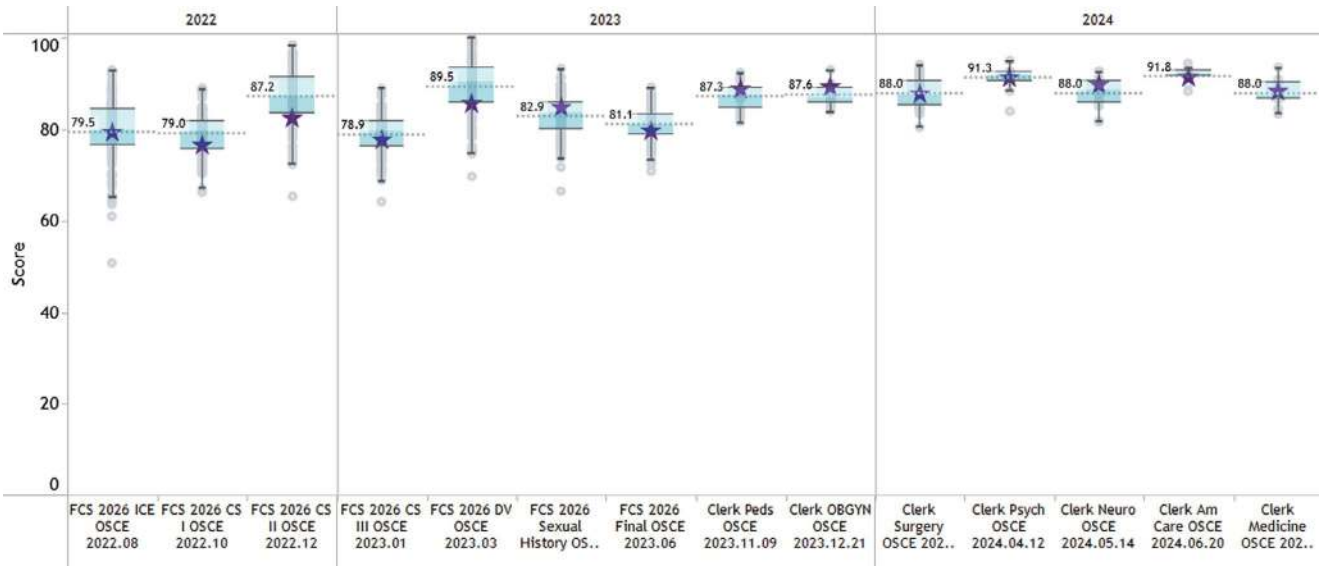


Fig. 6.16 Longitudinal views of communication skills over time This student (scores represented by a purple star) performed below the mean periodically throughout their first year of medical school but then, starting with clerkships, generally performed at or above the mean. A student viewing this data could see their progress over time, consider which specific communication skills have improved, and reflect on OSCEs where their scores were higher or lower than expected

mative assessment's reliability and validity (their consistency and accuracy) since they form the basis for conclusions about learners' abilities that often have significant consequences.

Below, we provide some examples of reporting on summative assessments for individual learners. Figure 6.15 shows elements of the customized report students receive as part of a comprehensive clinical skills OSCE.

Figure 6.16 depicts a dashboard view available to learners to view their summative assessment data over time.

Reporting for Educators/Coaches

The key difference between reporting for learners and reporting for educators/coaches is that educators and coaches are viewing individual learners' performance in the context of a group of learners and have a responsibility both to the learner and the cohort of learners.

If you, as an educator, have designed an OSCE to provide learners with opportunities to practice and hone their synthetic clinical skills across a variety of clinical scenarios and receive feedback on those skills for ongoing development and mastery, then the key questions you want to answer when reviewing the results are things like:

- Are there any learners we should be concerned about because their performance suggests they will require remediation or other efforts to ensure they are ready for the next stage of training?
- As we shift to competency-based medical education, the flip side of this question is newly important: are there learners demonstrating levels of mastery that suggest they are ready for the next stage of training? That is, could they begin that stage now, rather than waiting for the time-delineated stage to be complete?
- Can we identify patterns of performance that highlight general developmental and/or curricular effects (that is, expected strengths, weaknesses, and changes over time that align with anticipated levels and kinds of competence and with the design of the curriculum)?
- Can we identify individual deviations from those general patterns that illustrate more idiosyncratic journeys to competence? Can we still have confidence that competence will be achieved when needed or should we implement interventions to support these learners?
- How do I help an individual learner understand their performance within this larger context? What messages should they take from the results, what are the expectations, and what should they do if they don't meet (or do meet or exceed) the standards?

Finally, aside from high-stakes summative "pass/fail" exams where the OSCE and its assessments are sufficiently rigorous, reliable, valid, and consistent to permit a single judgement of competence, individual and group competence is almost always best understood from a longitudinal, developmental perspective. Low (or below expected) performance in one OSCE or case that then improves over time conveys a very different message, set of implications, and learning plans than continued low performance.

In Figs. 6.17, 6.18, and 6.19, we provide examples of data visualizations that support educators and coaches in answering these questions in collaboration with their learners.

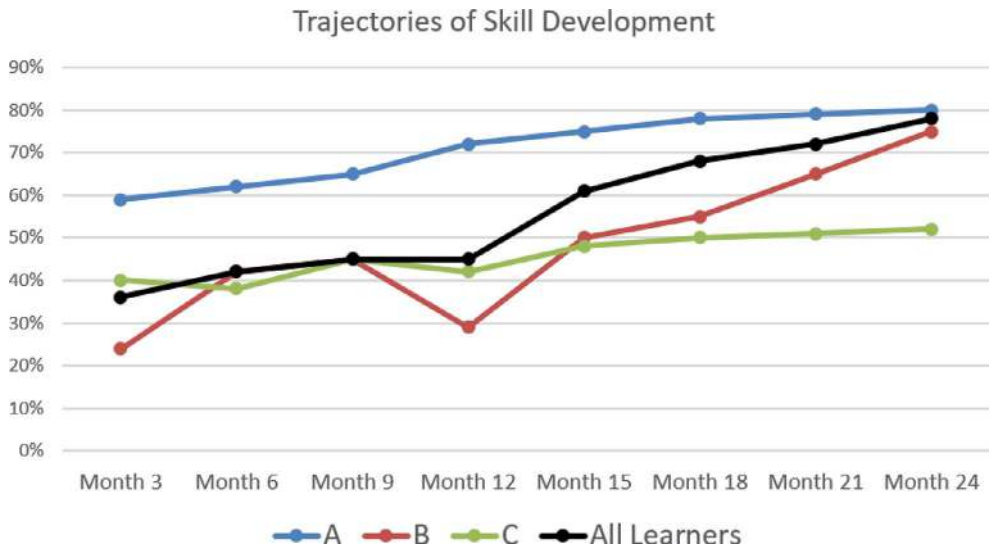


Fig. 6.17 Longitudinal (developmental) communication performance: patterns, interpretations, and implications. All learners (the black trend line) show a gradual increase in assessed performance of skills over time (based on mean values). Individual learners (A, B, and C) demonstrate different patterns or trajectories. Learner A: started out a bit ahead of all other learners and improved steadily but not as dramatically as the cohort average and ended up, at Month 24, at about the same level as the cohort. Learner B: started out a bit below the cohort average; started to "catch up"; then experienced a sharper decline in Month 12 (perhaps needing some time to consolidate skills); and finally also arrived, over the next year, at the same level as the cohort. Learner C: started closely with the rest of the cohort but then, instead of progressing rapidly over the next year, stayed static in their level of performance. These different profiles suggest different interventions—particularly Learner C

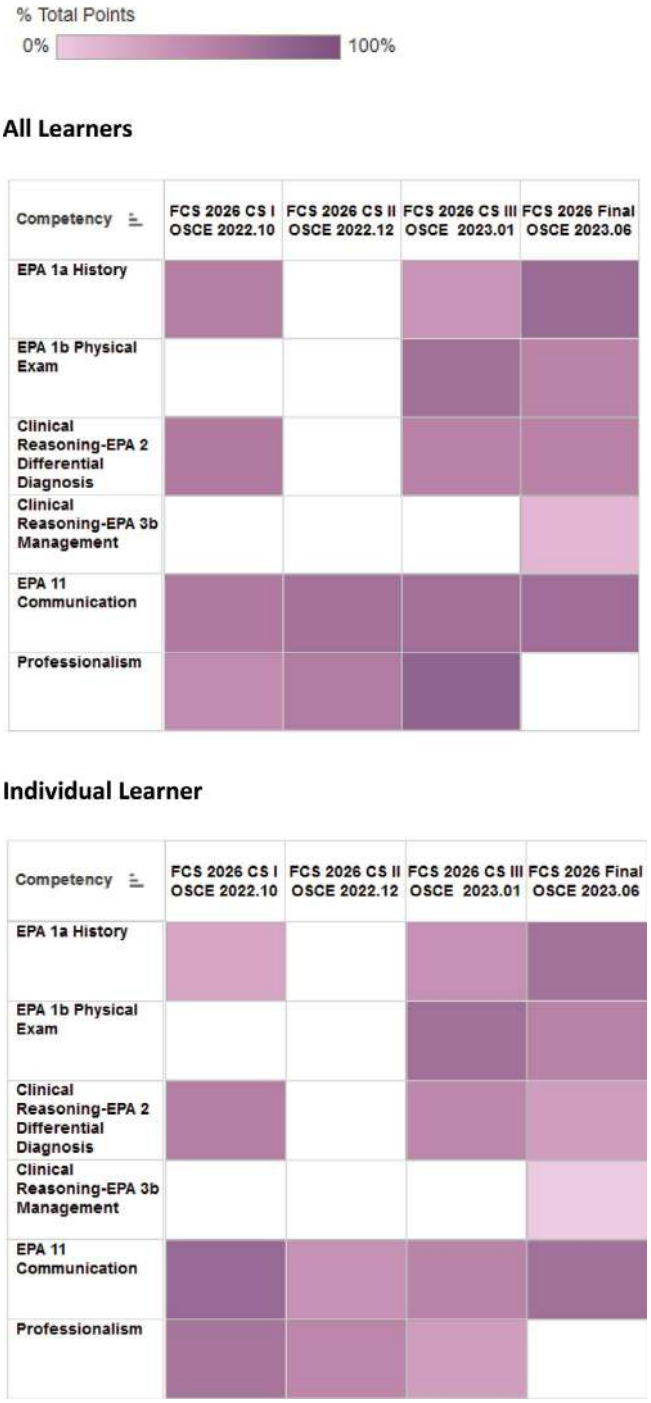


Fig. 6.18 Interpreting individual strengths and weaknesses in the context of cohort patterns. This "heat map" shows the % of total points awarded to the learners across competency areas over time, with the deeper the shade corresponding to higher percent. The top visualization describes all learners' scores by competency across four OSCEs arranged chronologically; the bottom visualization shows an individual learner's scores. Being able to compare the individual learner to all learners over time can help put an individual learner's patterns in context. The low score (indicated by the light pink shade) in the fourth OSCE for Clinical Reasoning-EPA 3b Management for the individual learner might be concerning if we weren't able to see that all learners received (on average) a similar low score. Conversely, while all learners appear to be improving in terms of their assessed professionalism, the individual learner is demonstrating a decrease in professionalism over the same time. Finally, the individual learner demonstrates strengths in EPA 1a History Gathering and EPA 11 Communication Skills in later OSCEs. This same pattern is seen for all learners. However, the individual students' Clinical Reasoning-EPA 1 Differential Diagnosis is declining over time and on the low side in the fourth OSCE, which is opposite the trend of the cohort, suggesting that this learner may need some additional support in developing their differential diagnosis skills

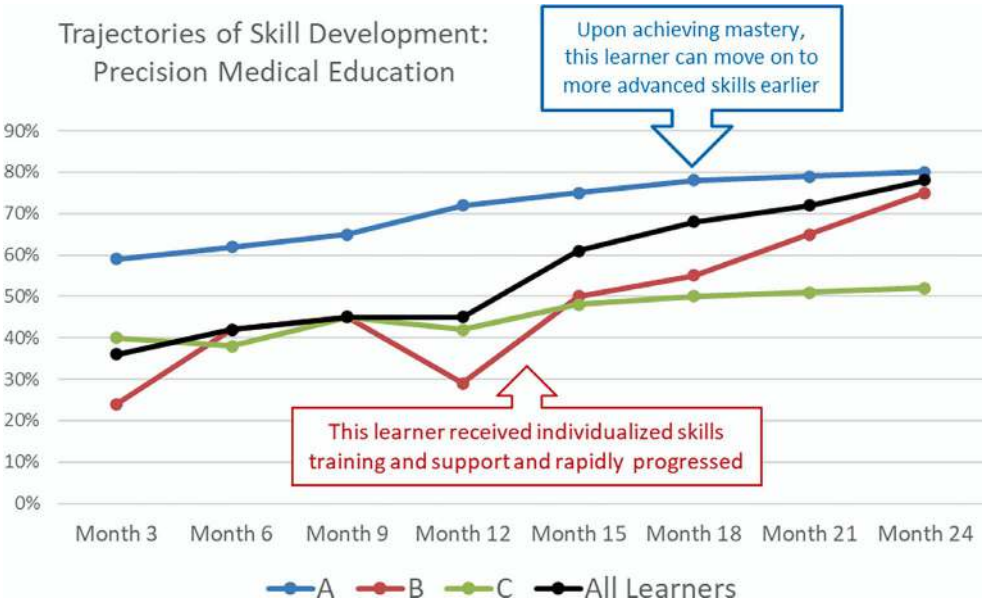


Fig. 6.19 Using outcomes to drive individualized learning (precision medical education). Example of using OSCE assessments to provide individualized training and support—meeting learners where they are

Reporting for Program/Curricular Evaluation

OSCE data, if collected within well-designed OSCEs with credible quality assurance and convincing psychometric properties, provide excellent outcomes for evaluating the degree to which programs/curriculum are achieving their intended objectives. While actual performance in clinical settings and the effectiveness of real-world skills in terms of quality of patient care is the pinnacle of outcome evidence, it is often very difficult to collect such assessments and even more difficult to do so in ways that are reliable and valid. OSCEs provide a very nice balance of feasibility, standardization, control, and compelling near real-world contexts and clinical challenges—such balance is not easily found in the world of assessment and explains the power of OSCEs in medical education. Thus, OSCE data tells us a great deal about learners' ability to demonstrate essential skills in controlled but realistic situations, representing a key method of not only judging individual competence (as described above) but also of evaluating the effectiveness of training programs and curricular efforts.

Using OSCE data for outcome evaluation shifts the focus to aggregate performance (with attention to variation) across cohorts of learners defined by exposure to educational interventions.

Cohorts can be defined based on the following:

- Admissions Cycles—In medical school, we examine the performance of medical school classes; in residency, cohorts are defined by PGY (postgraduate year). This determines not only the curriculum students undergo but also their scope of practice and degree of supervision.
- Time—Represented most often in terms of secular trends such as students entering medical school after the COVID-19 pandemic or simply over time.
- Curricular Exposure—Cohorts may experience different curriculum either by design or by happenstance; such curricular effects can be comprehensive (an entirely revised curriculum) or partial (the introduction of a new course or clinical training experience).
- Controlled (or Quasi-) Experiments—Cohorts defined through research design, such that learners could be randomly assigned to either of two curricular interventions or by taking advantage of a natural experiment wherein one cohort of learners is scheduled to participate in a newly updated curricular block while the previous cohort completed the unchanged curricular block.
- Individual Attributes—Learners may differ in factors such as prior experience, exposure, training programs, or educational experiences.

Program evaluation can use the OSCE-assessed outcomes associated with cohorts—defined in multiple ways above—to provide insight into the effectiveness of the educational programs.

Below, we describe some examples of this in practice to provide a sense of the ways in which OSCE data can be used for evaluation:

1. ***Evaluation of a New Approach to Teaching***: Pain management in medical school—Post-test only comparison of cohorts [27]:

   OSCE data were used to compare the pain assessment, management practices, and skills of a cohort of medical students who participated in a new curriculum on pain

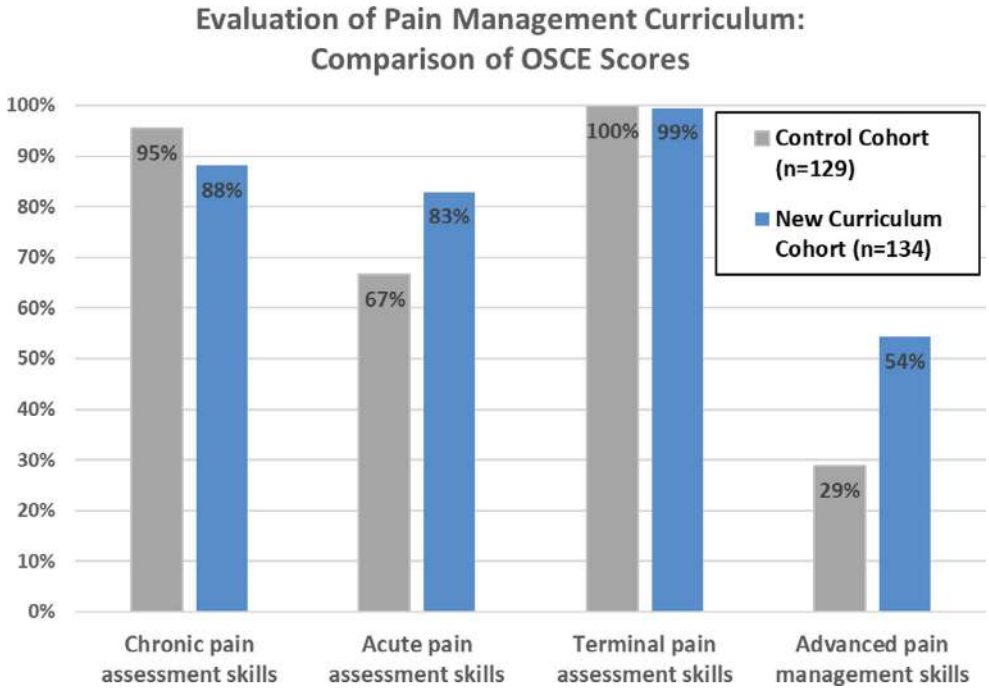


Fig. 6.20 Evaluation of pain assessment and management curriculum using OSCE data. Mean scores = % partly or well done adjusted for the following covariates: USMLE Clinical Knowledge Score, Neurology Shelf Exam Score, Overall OSCE Communication Score, and Overall OSCE History Gathering Score. All differences except for "Terminal pain assessment skills" were significant ($p < 0.05$)

management with those from the prior year who did not have the new curriculum. Three cases—abdominal pain (chronic), chest pain (acute), and shoulder pain (resulting from cancer)—were part of the routine OSCE program. Each had specific pain-related assessment items, including basic skills, such as obtaining a detailed description of the pain, and more advanced skills, such as counseling about pain medication and management options. We compared the distribution of ratings for the specific pain items between the cohorts and then also computed scores for domains including assessment and management (Fig. 6.20). This study found that the cohort that participated in the new curriculum had significantly better advanced pain management skills than the prior cohort.

2. ***Evaluation of an Initiative to Enhance Response to Social Determinants of Health (SDOH)****:* Unannounced standardized patient (USP) data on how primary care teams responded to SDOH after participating in audit/feedback sessions that provided teams with data on initial gaps in SDOH practices [28]:

   In this study, we used USPs to collect initial data on whether primary care teams elicited SDOH information from patients and, if elicited, how they responded to those social needs. These data were then fed back to the teams in the form of audit/feedback reports (Fig. 6.21). USPs were then deployed again to assess whether teams improved. As shown in Fig. 6.22, the study found that providing feedback did appear to be associated with increased rates of eliciting and responding to some social determinants of health.

3. ***Using Dynamic and Interactive OSCE Dashboards for Ongoing Evaluation of Medical School Curriculum***

   At NYUGSOM, the development of our OSCE dashboard has allowed us to review performance across the more than 40 OSCE cases that each student completes by graduation. We use this dashboard to revise and recalibrate our medical school's program of assessment based on review of OSCE data on core clinical skills across the curriculum. These data are used to determine whether students are achieving expected levels of skill development and competence at various stages throughout the curriculum and updating the curriculum in response. Ongoing review of these data has helped faculty ensure a comprehensive program of assessment that meets learners' needs for practice and feedback throughout the curriculum and make judgements about the readiness of graduates for residency.

   Access to OSCE data for many cohorts of students led to the recognition that students' communication skills were quite high (Fig. 6.23, view A) and therefore faculty felt comfortable with increasing the difficulty of some cases later in the curriculum to prepare graduates for the complexity of actual practice. However, review of physical examination skills identified specific areas where students needed more training and more opportunities to practice and master key maneuvers (Fig. 6.23, view B). Identification of variation in clinical reasoning skills within cohorts of students experiencing the same curriculum (Fig. 6.23, view C) spurred faculty in the preclerkship curriculum to provide more structured clinical

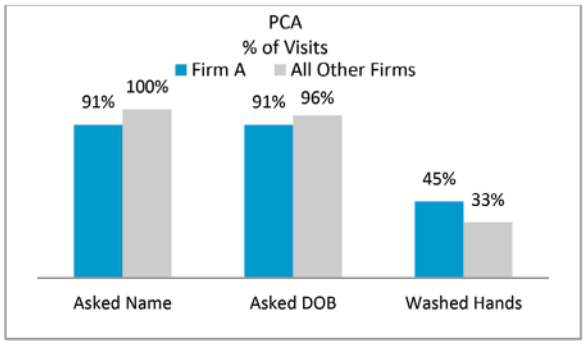
FIRM A Report

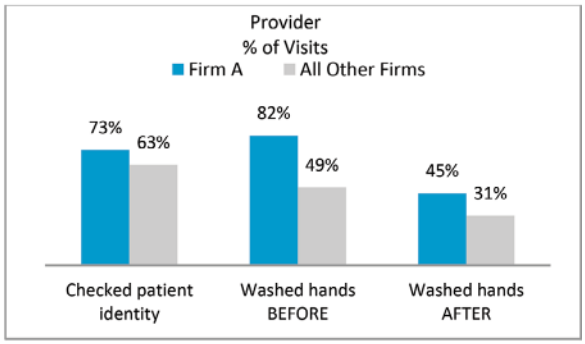
Feedback on safety, patient centeredness, quality of care, teamwork and response to social determinants of health for ***your team (Firm A)*** based on Unannounced Standardized Patient visits.

Timeframe: Apr – Sep 2018
FIRM A: 11 Visits
All Other Teams: 35 Visits

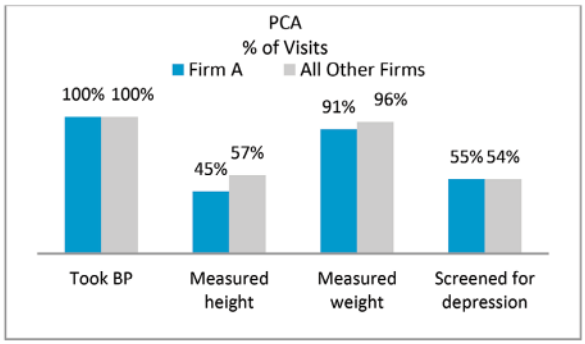
Includes data for the PCA (Patient Care Associate), Provider, and the Team.

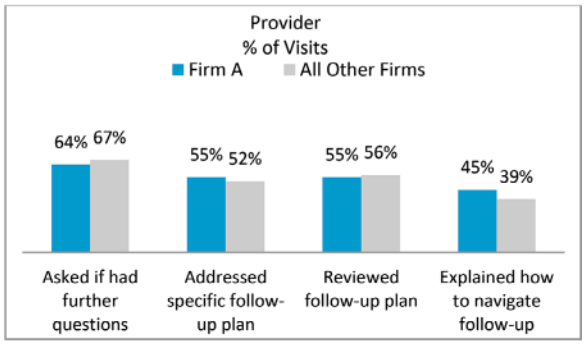
SAFETY





QUALITY OF CARE





OVERALL TEAMWORK

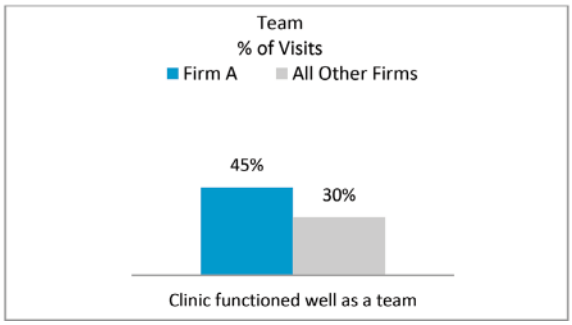


Fig. 6.21 Clinical team report (Firm A) based on unannounced standardized patient data (audit/feedback report)

SOCIAL DETERMINANTS OF HEALTH REPORT

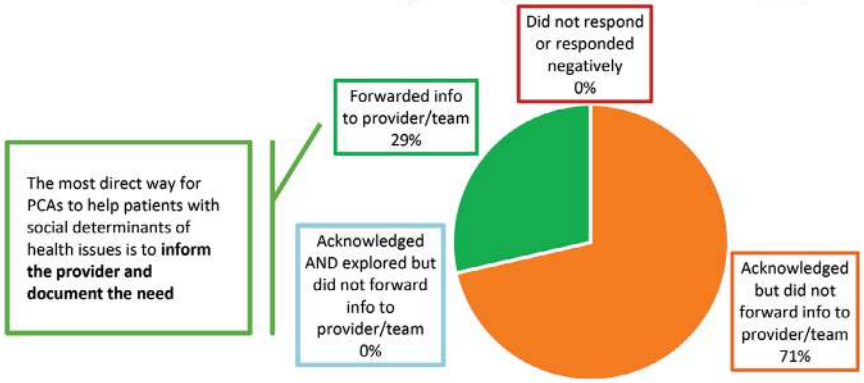
Unannounced Standardized Patients visit the clinic periodically and are trained to ***volunteer specific housing, financial/economic, and social information that is relevant to their condition/care.***

These same patients also have ***underlying*** housing, financial/economic and social issues that they only bring up if asked about generally (e.g., "tell me about your housing situation?").

We track what happened in these visits to ***understand the degree to which teams find out about these issues, respond to them (by referring to services), and record them in the health record.*** **Below we share your team's data from the past 11 months.**

PCA

When the Standardized Patient *volunteered* housing, financial, or social information to the PCAs on Firm A, this is what happened (% of visits):



Good Job!

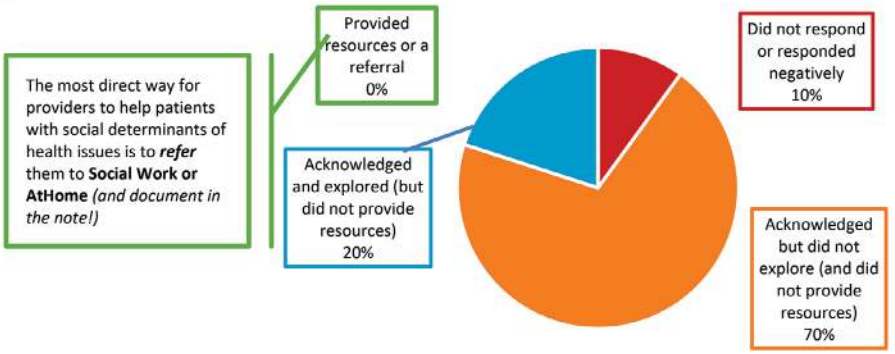
COMPARED TO THE PREVIOUS REPORT

PCAs did not respond or responded negatively in 0% of visits compared to 17% in the last report.

PCAs acknowledged more often in this period (71%) compared to the last (33%).

PROVIDER

When the Standardized Patient *volunteered* housing, financial, or social information to the provider on Firm A, this is what happened (% of visits):



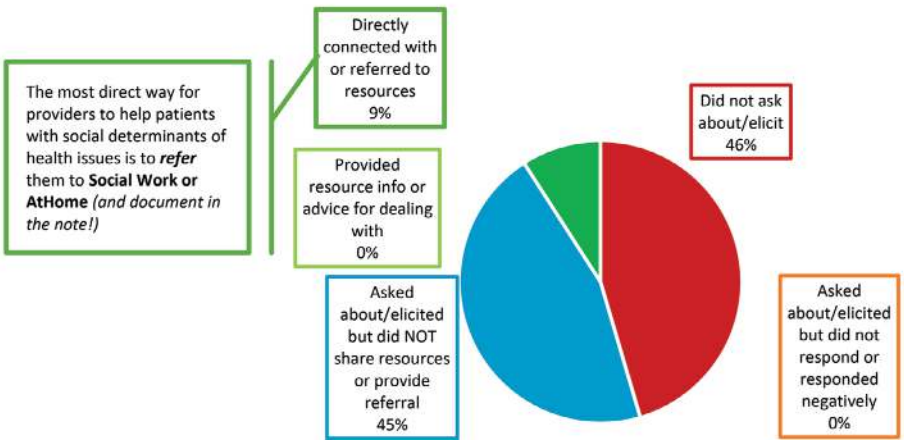
Acknowledging SDoH is not enough

COMPARED TO PREVIOUS REPORT

Providers ***acknowledged but did not explore*** in less visits (70% now vs. 75%) but ***acknowledged AND explored*** at a higher rate (20% vs. 13%).

Providers also provided resources or referrals in ***fewer*** visits (0% vs 13%).

When the Standardized Patient's social, financial, or housing information *needed to be elicited* by the provider



Good Job!

COMPARED TO PREVIOUS REPORT

Providers ***did not respond or responded negatively*** in less visits (0% vs 11%).

But...

Providers directly connected or referred the patient to resources in less visits (9% vs. 22%).

Fig. 6.21 (continued)

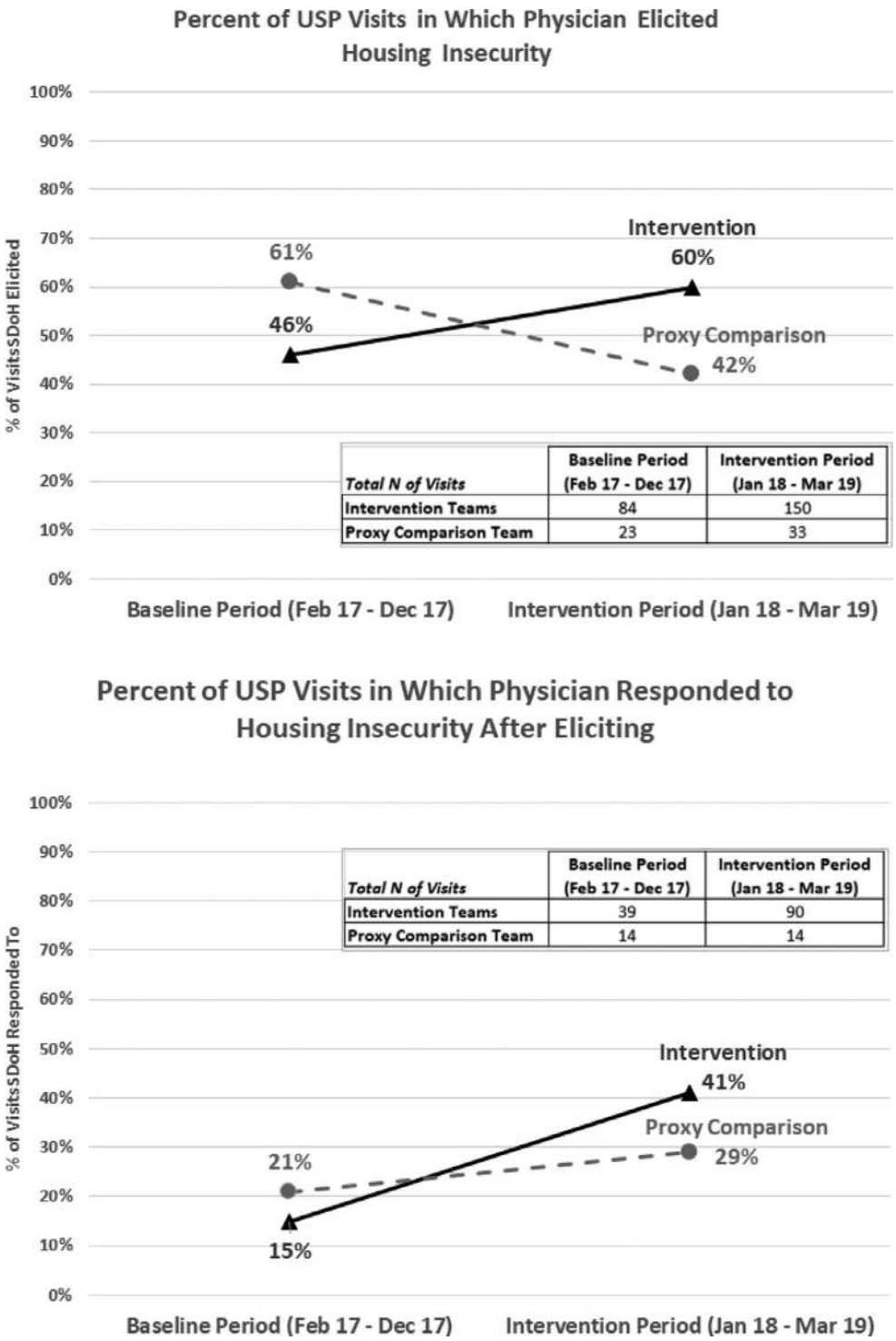


Fig. 6.22 Impact of audit/feedback on response to social determinants of health. Charts above describe the percent of unannounced standardized patient visits in which the physician elicited an issue of underlying housing insecurity from the USP (top) and then responded to that issue (bottom) for teams that had received audit/feedback reports from prior USP visits with social determinant of health issues compared to a proxy comparison team that had not received audit/feedback reports [28]

reasoning curriculum in the form of illness scripts and similar approaches that could be used by learners across the spectrum of clinical reasoning development. This also prompted assessors to provide more granular feedback on defined elements of clinical reasoning demonstrated in students' written notes.

Our medical school's educational program objectives are defined as the core competencies associated with the Physicians' Reference Competency Set (PCRS), and we use OSCE data at the cohort level, in combination with direct observation data and other outcomes, to determine if the medical school has achieved these educational program objectives.

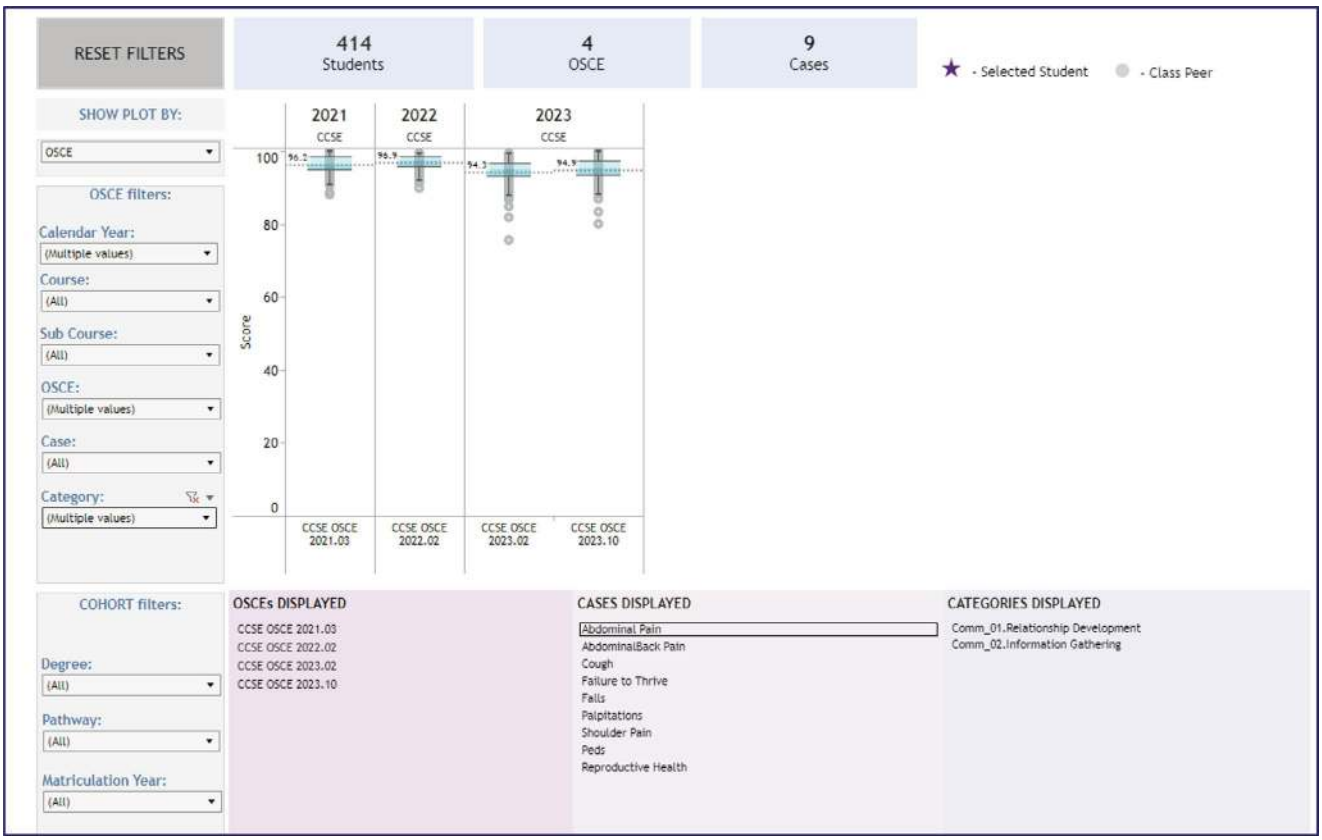


Fig. 6.23 OSCE dashboard views for curricular evaluation and planning

View A: Communication Skills. Assessment of OSCE performance data for many cohorts of students led to the recognition that students' communication skills were quite high and therefore faculty felt comfortable increasing the difficulty of some cases later in the curriculum, to prepare graduates for the complexity of actual practice

View B (below): Physical Exam Skills. NOTES: Review of physical examination skills identified specific areas where students needed more training and more opportunities to practice and master key maneuvers

View C (below): Clinical Reasoning Skills

Notes: Identification of variation in clinical reasoning skills within cohorts of students experiencing the same curriculum spurred faculty in the pre-clerkship curriculum to provide more structured clinical reasoning curriculum in the form of illness scripts and similar approaches that could be used by learners across the spectrum of clinical reasoning development and more granular feedback from assessors on defined elements of clinical reasoning demonstrated in students' written notes. This led to an increase in clinical reasoning skills for that cohort of students

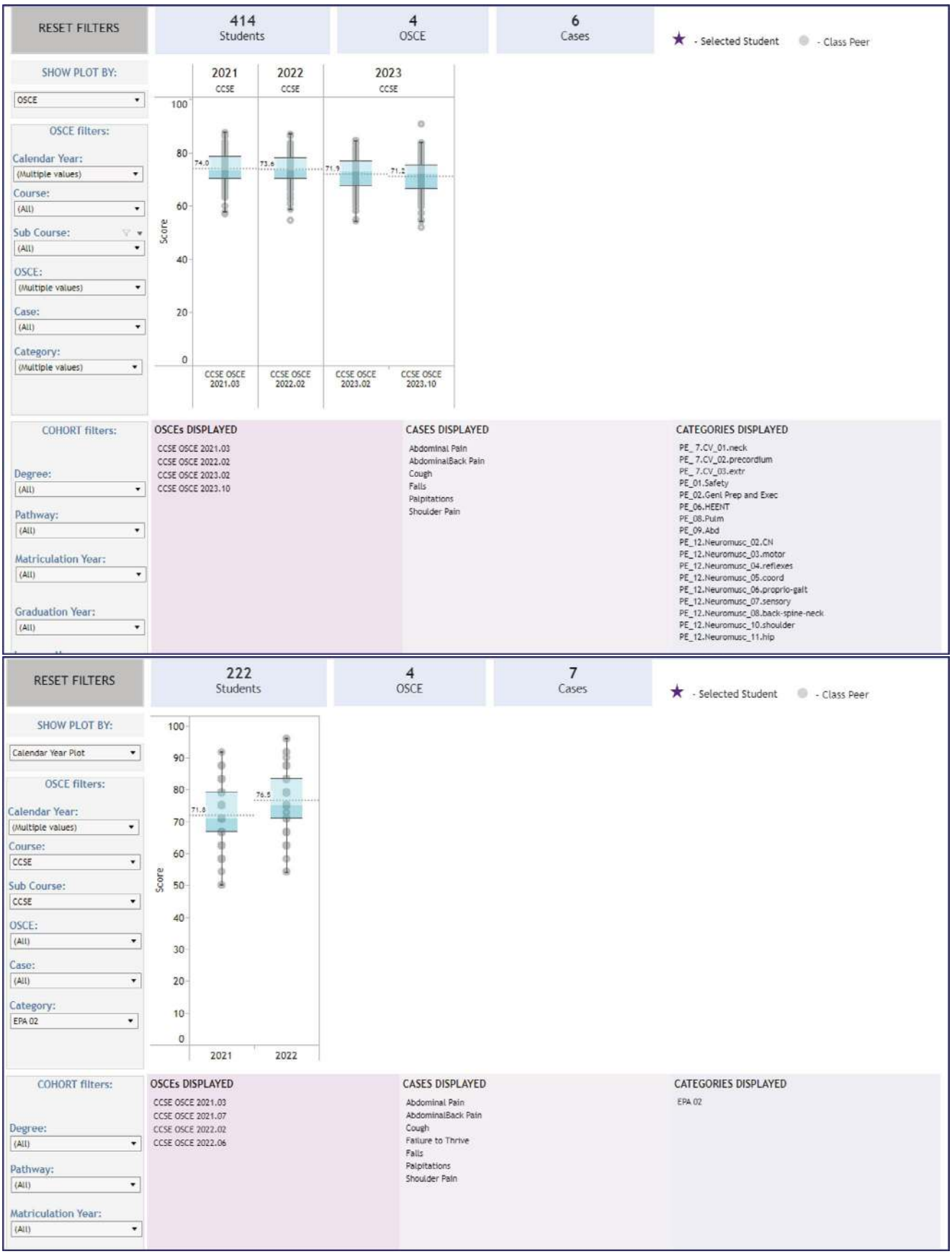


Fig. 6.23 (continued)

4. ***Evaluating the Impact of OSCE Experiences***: Understanding how OSCEs may influence future competence [29]:

   This qualitative research study evaluated the ways in which learners' participation in OSCE cases related to transgender health may influence their performance farther along in their training and found that such formative experience can activate learners to develop targeted skills, seek out practice opportunities, and build their knowledge base for the future.

5. ***Evaluating Trends over Time:*** Communication skills over time in medical school provide evidence for increased emphasis on education and counseling skills [30]:

   In this study, we tracked individual students' communication skills performance on three core OSCEs throughout medical school and found that while communication skills overall improved (in aggregate) from the start of medical school to the end of the core "doctoring" course (end of second year), there was substantial decline in the domain of education and counseling skills from then to a high-stakes OSCE completed at the end of the third year (Fig. 6.24). Review of the sub-domains of communication show that education and counseling communication skills improved the most from the start of the first year to the end of the first year but then declined for the third-year high-stakes OSCE, suggesting a need for continued focus on this important skill.

6. ***Evaluating the Impact of Curricular Revision on Students' Communication Skills***: Using OSCE data to compare the communication skills of students exposed to a newly revised curriculum (C21 Personalized Pathways) with those of the prior cohort of students [31]:

   Our medical school implemented a new curriculum that focused on providing students with more personal-

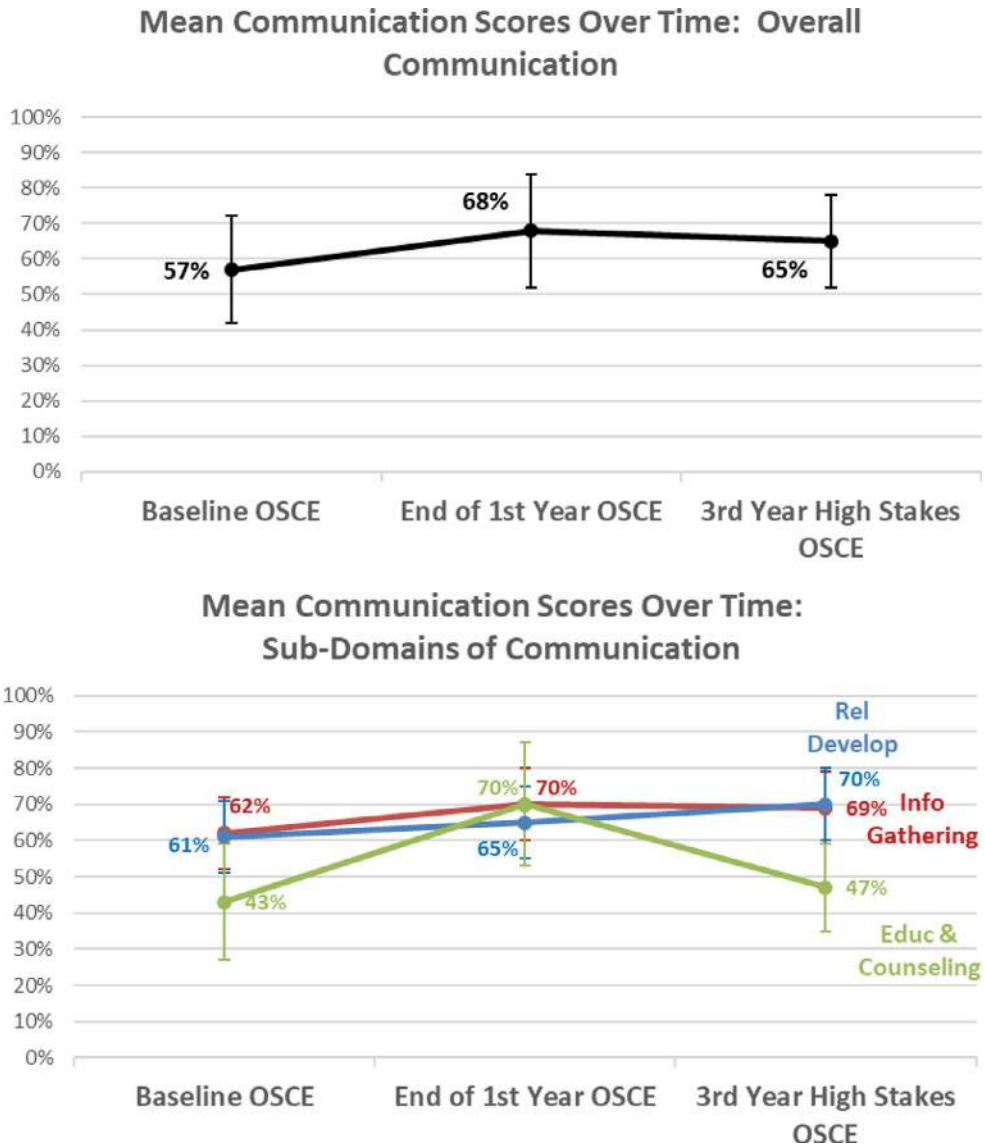


Fig. 6.24 OSCE communication skills over time: baseline to end of first year OSCE to third year high-stakes OSCE. Evaluation of our communication skills curricula over time [30] shows that medical students improve significantly in their overall communication skill throughout their first 2 years of medical school. Review of the sub-domains of communication we measured shows that education and counseling communication skills improved the most from the start of the first year to the end of the first year but then declined for the third-year high-stakes OSCE, suggesting a need for continued focus on this important skill

ized pathways through medical school, including shortening the duration of the pre-clerkship stage of the curriculum from about 15 months to about 11 months. Critical to the success of this revision was ensuring that students would still be ready for the clerkship stage in terms of their core clinical skills. To that end, we compared the communication skills of students as assessed in the end of clinical skills course "final" OSCE between the two curriculum cohorts: the class that completed the 15-month version of the course vs. the class that completed the accelerated 11-month version (Fig. 6.25). In addition, because we have a longitudinal program of integrated clinical skills assessment, we were able to control for the cohorts' baseline communication skills by including those assessments from the Introductory Clinical Encounter OSCE—a formative three-station OSCE conducted in the first few weeks of medical school—in the statistical analyses. We found that the communication skills of students in both cohorts were very similar at the end of the course and that this was true within the three specific domains of communication assessed (information gathering, relationship development, education, and counseling) when controlling for initial skill levels. No significant differences were found suggesting that the acceleration had not adversely affected students' preparation for clerkships when it came to communication skills.

7. ***Evaluating Whether Core Clinical Skills Are Affected by Completing Medical School as Part of an Accelerated Pathway:*** Comparing performance on a high-stakes OSCE between 3-year and 4-year pathway students [32]:

   In order to ensure that students who participated in our medical school's accelerated 3-year pathway were as ready for residency as students in the traditional 4-year program, we compared the performance of these two cohorts on our Comprehensive Clinical Skills Exam—a high-stakes, rigorous, post-clerkship OSCE designed to make summative judgments about students' development of clinical competency. As visualized in Fig. 6.26, there were no significant differences between the 3-year and 4-year pathway students in terms of their core clinical skills mean scores, indicating that both pathways are acceptable preparation for residency.

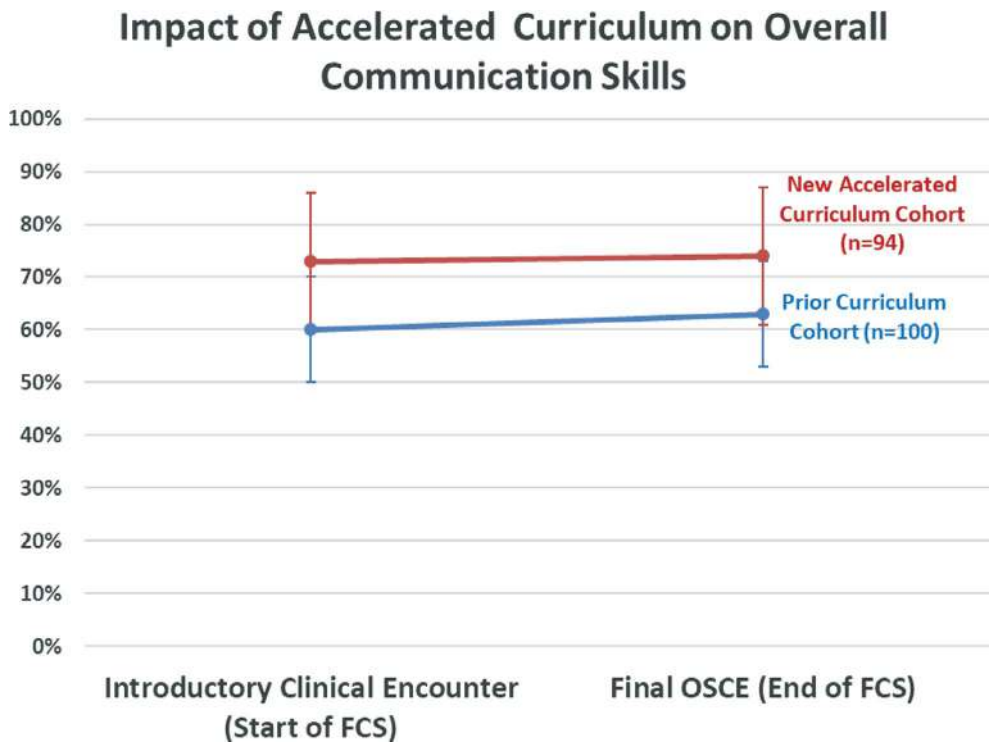


Fig. 6.25 Impact of new accelerated curriculum on communication skills. A new accelerated curriculum was introduced at our medical school and led to the core skills of doctoring first year course (*Foundational Clinical Skills (FCS)) being shortened from about 16 months to about 11 months. The impact of that acceleration was monitored by comparing the overall communication skills of the cohort experiencing the new accelerated curriculum at the end of the course with those from the prior cohort while controlling for the cohorts' initial communication skills (using an early OSCE). No significant differences were found suggesting that the acceleration had not adversely affected students' preparation for clerkships when it came to communication skills

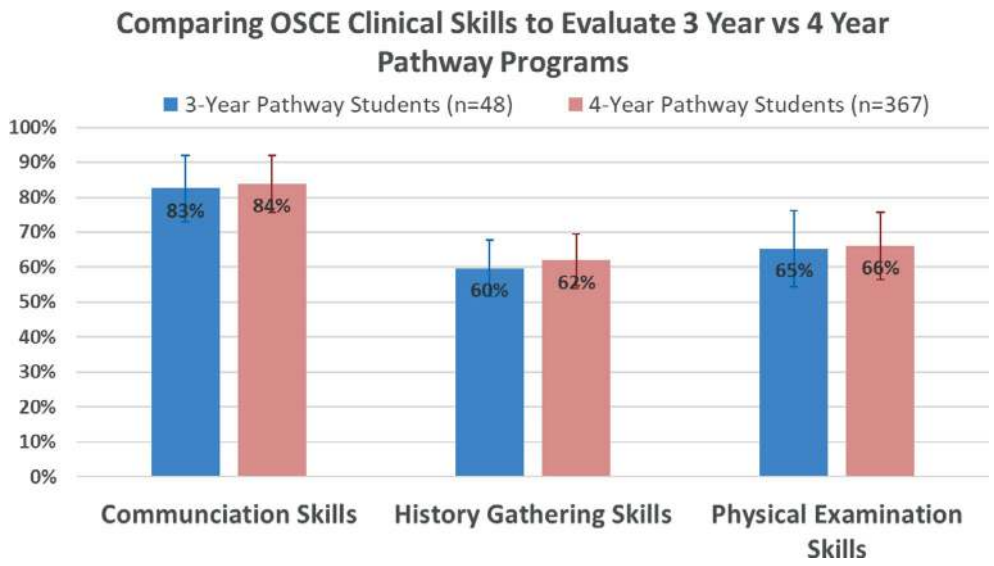


Fig. 6.26 Comprehensive performance-based assessment to evaluate 3-year vs. 4-year pathway program. In order to ensure that students who participated in our medical school's accelerated, 3-year pathway were as ready for residency as students in the traditional 4-year program, we compared the performance of these two cohorts on our Comprehensive Clinical Skills Exam—a high-stakes, rigorous, post-clerkship OSCE designed to make summary judgments about students' development of clinical competency. As visualized above, there were no significant differences between the 3-year and 4-year pathway students in terms of their core clinical skills mean scores

Advanced Techniques for Analyzing OSCE Data

This chapter has focused on fairly simple, widely accessible methods for describing, analyzing, and visualizing OSCE data. There are many more advanced approaches to understanding OSCE data, and we very briefly and generally describe a few of them here. The power of these approaches lies in their ability to reflect the complexity of OSCE data as multidimensional, with multiple sources of variation, potential measurement error, and confounding, with varying implications for education and training.

Patterns/Profiles (Cluster Analysis, Latent Profile/Class Analysis)

As noted throughout this chapter, scores are often computed that summarize assessment items by domain or across cases or across and entire OSCE. These summary scores are very useful for providing synthetic understanding of performance, but they do reduce the complexity of the data and may therefore fail to capture important patterns of results.

One way of exploring the pattern of results in OSCE data is through cluster analysis. Cluster analysis is an exploratory data analysis approach that sorts observations (data or cases) into groups (clusters) [33]. The clusters are defined empirically, through an analysis of the data, and are not based on prior knowledge of which cases "should" fall into which clusters or even what clusters are likely to exist. Instead, the clustering algorithm (and there are many approaches to clustering) maximizes the similarity of scores within a cluster while simultaneously maximizing the dissimilarity in scores between clusters. Cluster analysis can identify groups of students who perform similarly (and differently) across a range of OSCE-assessed domains—patterns that might be obscured by looking only at mean values. Such clusters might then suggest different curricular approaches.

In addition, summary scores can mask deficits in particular areas if offset by strengths in other areas. Latent profile analysis (LPA) is a multivariable statistical analysis technique that clusters individuals based on item response patterns of continuous data (see Becker-Haimes et al. [34] for an example from healthcare). If using categorical/ordinal data, latent class analysis is used. As seen in Fig. 6.27, we used LPA to cluster students by item performance patterns to identify and describe subsets of learners with similar strengths and weaknesses using OSCE communication skill assessments. With nine cohorts of student OSCE data, we used LPA to understand communication skills development at the individual item level [35]. Three specific communication skill items—"asked questions to see what you understood," "allowed you to talk without interrupting," and "nonverbal behavior enriched communication"—appeared to discriminate among three clusters of students: generally high-performing students, average, and lower-performing students.

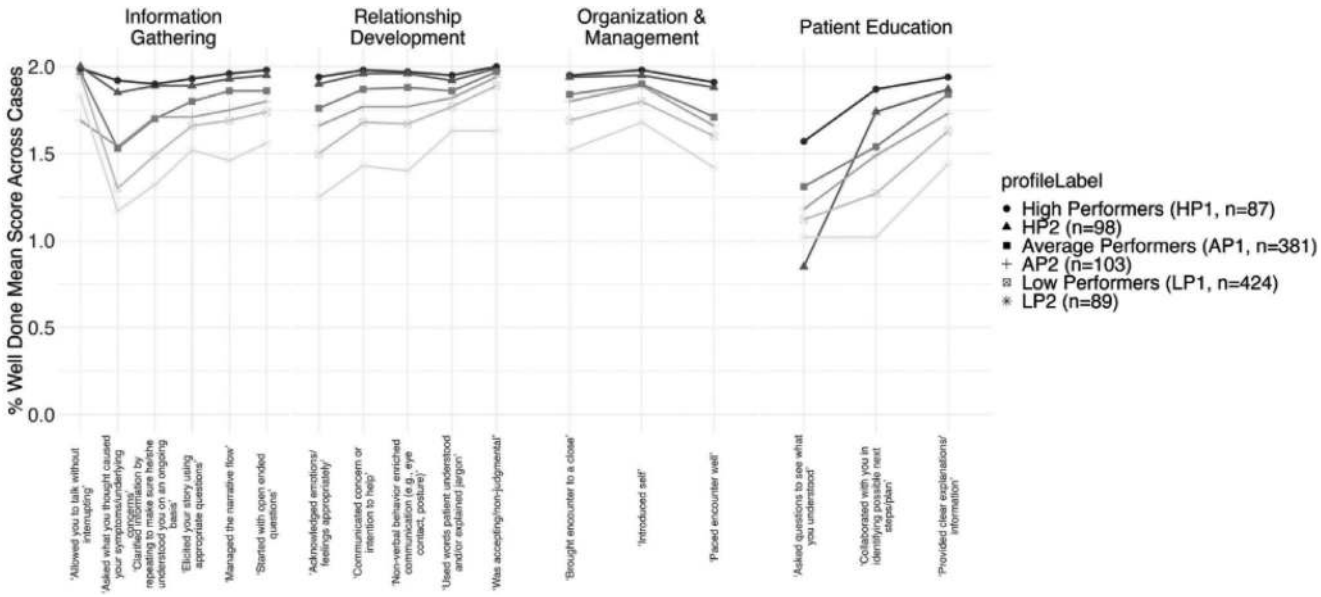


Fig. 6.27 Understanding medical student paths to communication skills expertise using latent profile analysis

Multilevel Modeling

OSCE data are inherently nested data—meaning that skills are assessed within cases, assessments may be completed by different raters, and cases are nested within OSCEs. In addition, characteristics of the learners may interact with or influence performance, and characteristics of the context (settings, situations, timing, etc.) may also interact with or influence performance. This makes OSCEs well suited for multilevel modeling approaches or statistical models that include variation at multiple levels. Multilevel models can take many different forms and may be known by any of the following labels: hierarchical linear models, linear mixed-effect model, mixed models, nested data models, random coefficient, random-effects models, random parameter models, or split-plot designs. The units of analysis are usually individuals (at a lower level) who are nested within contextual/aggregate units (at a higher level). While the lowest level of data in multilevel models is usually an individual, repeated measurements of individuals may also be examined.

Within an OSCE, a multilevel model might seek to describe how performance is influenced by case and also by the SP playing the cases and conducting the assessments. Similarly, a multilevel model might seek to account for the effects of the fact that students completing an OSCE were educated and trained in small groups that might influence their performance. Such models could also seek to explore whether distinct groups of students—as defined by sociodemographic characteristics—might have different patterns of performance across different kinds of cases or OSCEs or in different settings (OSCEs vs. real world practice as assessed by USPs).

Generalizability Theory (G Theory)

Generalizability theory is an approach to maximizing reliability by identifying sources of variance in scores in order to minimize measurement error. In G theory [36], sources of variation are referred to as *facets*. Facets are the parameters, groups, or conditions that affect assessments, including, as described above, the assessed individuals, raters, item or assessment characteristics, time, context, or setting. In a perfect world, one could control all of these influences and therefore isolate the true signal from the noise. The goal of G theory is to view all of these facets (and their interactions) as potential sources of error and to quantify the amount of error caused by each. The G in G Theory comes from the idea that when conducting assessments, one hopes to use the results to generalize more broadly—to the competence of the individual (e.g., from a small number of items to what would be found with a larger number of items), to a future point in time, to other types of learners, to a different setting, etc.

G studies are designed to answer specific questions about which sources of variance matter. First, one has to decide which facet will serve as the object of measurement (e.g., the systematic source of variance). The remaining facets of interest are then considered to be sources of measurement error. In most cases, the object of measurement will be the person to whom a score is assigned. Ideally, most of the measured variance will be attributed to individual differences, with only a negligible amount of variance attributed to the remaining facets (e.g., rater, time, setting).

In one example from the literature, G Theory was used to determine the consistency of assessment scores in a post-clerkship neurology OSCE delineated by the following

facets: cases, domains, and items [37]. They found that the majority of variation was attributable to differences by domain, suggesting that individual students are likely to have distinct strengths and weaknesses. In another study [8], G Theory was used to determine that the major sources of variance in a fifth-year medical student 18-station OSCE were stations/cases and students, while sites (setting) and versions of the assessments had minimum variance. This supports the use of the OSCE for assessing individual students' competence but also highlights that such competence can vary (as it did in this study) substantially across stations/cases (content-specificity).

Growth Trajectories and Predictive Analytics

At its most simple, growth trajectories describe patterns over time, and predictive analytics can be defined as using data to predict future trends and events. In assessment terms, predictive analytics closely follows the concept of predictive validity—the ability of a measurement or assessment to predict future behavior—but is extended to include multiple inputs into predicting future performance. Think multilevel modeling meets predictive validity!

Predictive analytics help us answer the following question: what does what we know about performance in an OSCE tell us about future performance? This question is at the heart of much of what we care about in OSCE assessment as it informs how we interpret performance, what we do educationally or training-wise in response to OSCE data, and how we truly implement competency-based medical education—how do we know individuals are ready for the next stage? There are many, many ways to "do" predictive analytics, but below, we highlight three examples from the literature to give readers a sense of how predictive analytics can take us to the next level in terms of understanding and acting upon OSCE assessment data.

Following two cohorts of medical students, Martin and Jolly [38] assessed the degree to which performance in a third-year OSCE was associated with subsequent clerkship exams and found a moderate (and significant) association between the two. Poor performance in the OSCE was strongly associated with subsequent poor performance in the clerkship exams: students in the lowest three deciles of OSCE performance were six times more likely to fail a clerkship exam. Similar results were found in a study of the degree to which scores on an OSCE predicted national exam performance in a medicine residency program [39], with medium-sized associations found within each PGY. When all years were entered into the model, PGY2 and PGY4 years predicted failure on the national examination. Finally, a study of emergency medicine residents found that their performance on a five-station PGY1 OSCE at the start of training predicted future resident performance based on faculty assessments of the residents in the emergency department [40].

As our OSCE data accumulates over time, we are poised to be able to reap the benefits of these longitudinal databases to answer the next generation of predictive analytics questions: what are the key features (combinations of variables) including OSCE performance that predict future performance, and does OSCE performance predict both actual practice and patient outcomes?

Artificial Intelligence (AI): Machine Learning, Natural Language Processing (NLP), and Generative AI (genAI)

We cannot close out this chapter without acknowledging the tremendous promise genAI and large language models (LLMs) have for OSCEs [41]. While truly in its infancy, AI has many potential use cases in OSCEs. Such uses include examples of OSCE development, like generating cases and developing virtual SPs, that create low-stakes, cost-effective means to expand access to a wider range of training, practice, and feedback opportunities. From an assessment and reporting perspective, there is lots of interest and promising initial work in using genAI to analyze and categorize both quantitative and qualitative (written or verbal comments, feedback) into grading rubrics or milestones in order to move raw data to the next level of interpretability and action. See, for example, Jani and colleagues' [42] use of machine learning models to automatically label interview content and communication skills.

Given that assessing clinical reasoning and the quality of documentation in patient notes written as part of OSCEs is one of the most time- and resource-intensive of assessment elements in an OSCE, genAI holds particular promise. We, like many others, have piloted using genAI to grade and provide feedback on written notes to learners. In one recent example, we created a model note for the OSCE case and then prompted a genAI engine to provide feedback to students based on a comparison of their written note and this exemplar note. Some students found this feedback to be very helpful—grounded in concrete examples, personalized, actionable—while others were dismayed to receive feedback from an LLM.

More innovative approaches consider whether LLMs can identify patterns of communication, skills, and/or practices in OSCEs that have implications for subsequent effectiveness as a healthcare provider and could therefore be used to develop real-time observational systems to give feedback to healthcare providers in time to correct those practices. Optimizing these uses of AI could lead to dramatic improvements in the efficiency, speed, and accuracy of OSCE assessment, as well as the usefulness and interpretation of the results.

References

1. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, McLeod L, Delacqua D, Delacqua F, Kirby J, Duda SN. REDCap consortium, the REDCap consortium: building an international community of software partners. J Biomed Inform. 2019;95 https://doi.org/10.1016/j.jbi.2019.103208.
2. Borsboom D, Molenaar D. In: Wright JD, editor. Psychometrics in international encyclopedia of the social and behavioral sciences. 2nd ed. Elseveir; 2015. p. 418–22. https://doi.org/10.1016/B978-0-08-097086-8.43079-5. https://www.sciencedirect.com/science/article/pii/B9780080970868430795.
3. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. Med Educ. 2011;45(12):1181–9. https://doi.org/10.1111/j.1365-2923.2011.04075.x. Epub 2011 Oct 11
4. Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. Med Educ. 1988;22(4):325–34. https://doi.org/10.1111/j.1365-2923.1988.tb00761.x.
5. Messick S. Validity. In: Linn RL, editor. Educational measurement. 3rd ed. New York: Macmillan; 1989. p. 13–103.
6. Auewarakul C, Downing SM, Praditsuwan R, Jaturatamrong U. Item analysis to improve reliability for an internal medicine undergraduate OSCE. Adv Health Sci Educ Theory Pract. 2005;10(2):105–13. https://doi.org/10.1007/s10459-005-2315-3.
7. Lee M, Hernandez E, Brook R, Ha E, Harris C, Plesa M, Kahn D. Competency-based standard setting for a high-stakes Objective Structured Clinical Examination (OSCE): validity evidence. MedEdPublish (2016). 2018;7:200. https://doi.org/10.15694/mep.2018.0000200.1. PMID: 38074586; PMCID: PMC10701848.
8. Trejo-Mejía JA, Sánchez-Mendiola M, Méndez-Ramírez I, Martínez-González A. Reliability analysis of the objective structured clinical examination using generalizability theory. Med Educ Online. 2016;21:31650. https://doi.org/10.3402/meo.v21.31650. PMID: 27543188; PMCID: PMC4991996.
9. Patrício M, Julião M, Fareleira F, Young M, Norman G, Vaz Carneiro A. A comprehensive checklist for reporting the use of OSCEs. Med Teach. 2009;31(2):112–24. https://doi.org/10.1080/01421590802578277.
10. Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: review of metrics—AMEE guide no. 49. Med Teach. 2010;32:802–11. https://doi.org/10.3109/0142159X.2010.507716.
11. Revelle W, Condon DM. Reliability from α to ω: a tutorial. Psychol Assess. 2019;31(12):1395–411. https://doi.org/10.1037/pas0000754. Epub 2019 Aug 5
12. McNeish D. Thanks coefficient alpha, we'll take it from here. Psychol Methods. 2018;23(3):412–33. https://doi.org/10.1037/met0000144. Epub 2017 May 29. Medical Education 2016: 50: 351–358 10.1111/medu.12942.
13. Crutzen R, Peters GJY. Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. Health Psychol Rev. 2015. Advance online publication. https://doi-org.ezproxy.med.nyu.edu/10.1080/17437199.2015.1124240
14. Li CH. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. Behav Res Methods. 2016;48(3):936–49. https://doi.org/10.3758/s13428-015-0619-7.
15. Ark T, Kalet A, Tewksbury L, Altshuler L, Crowe R, Wilhite J, Hardowar K, Zabar S, Gillespie C. Validity evidence for the clinical communication skills assessment tool (CCSAT) from 9 years of implementation in a high stakes medical student OSCE. Patient Educ Couns. 2024;127:108323. https://doi.org/10.1016/j.pec.2024.108323. Epub 2024 Jun 2
16. Yousuf N, Violato C, Zuberi RW. Standard setting methods for pass/fail decisions on high-stakes objective structured clinical examinations: a validity study. Teach Learn Med. 2015;27(3):280–91. https://doi.org/10.1080/10401334.2015.1044749.
17. Danette WM, Norcini JJ. How to set standards on performance-based examinations: AMEE guide no. 85. Med Teach. 2014;36(2):97–110. https://doi.org/10.3109/0142159X.2013.853119.
18. Boursicot KAM, Roberts TE, Pell G. Standard setting for clinical competence at graduation from medical school: a comparison of passing scores across five medical schools. Adv Health Sci Educ Theory Pract. 2006;11:173–83. https://doi.org/10.1007/s10459-005-5291-8.
19. Downing SM, Tekian A, Yudkowsky R. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. Teach Learn Med. 2006;18:50–7.
20. Sargeant J, Armson H, Driessen E, et al. Evidence-informed facilitated feedback: the R2C2 feedback model. MedEdPORTAL. 2016;12:10387. https://doi.org/10.15766/mep_2374-8265.10387.
21. Watling CJ. Unfulfilled promise, untapped potential: feedback at the crossroads. Med Teach. 2014;36:692–7. https://doi.org/10.3109/0142159x.2014.889812.
22. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. Med Educ. 2019;53:76–85. https://doi.org/10.1111/medu.13645.
23. Sargeant J, Mann K, Manos S, et al. R2C2 in action: testing an evidence-based model to facilitate feedback and coaching in residency. J Grad Med Educ. 2017;9:165–70. https://doi.org/10.4300/jgme-d-19-00508.1.
24. Sargeant J, Lockyer JM, Mann K, et al. The R2C2 model in residency education: how does it foster coaching and promote feedback use? Acad Med. 2018;93:1055–63. https://doi.org/10.1097/acm.0000000000002131.
25. Lockyer J, Armson H, Konings KD, et al. In-the-moment feedback and coaching: improving R2C2 for a new context. J Grad Med Educ. 2020;12:27–35.
26. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. Med Educ. 2016;50:933–42. https://doi.org/10.1111/medu.1306.
27. Stevens DL, King D, Laponis R, Hanley K, Zabar S, Kalet A, Gillespie C. Medical students retain pain assessment and management skills long after an experiential curriculum: a controlled study. Pain. 2009;145(3):319–24.
28. Gillespie C, Wilhite JA, Hanley K, Hardowar K, Altshuler L, Fisher H, Porter B, Wallach A, Zabar S. Addressing social determinants of health in primary care: a quasi-experimental study using unannounced standardised patients to evaluate the impact of audit/feedback on physicians' rates of identifying and responding to social needs. BMJ Qual Saf. 2023;32(11):632–43. https://doi.org/10.1136/bmjqs-2021-013904. Epub 2022 May 27
29. Greene RE, Blasdel G, Cook TE, Gillespie C. How do OSCE cases activate learners about transgender health? Acad Med. 2020;95(12S Addressing Harmful Bias and Eliminating Discrimination in Health Professions Learning Environments):S156–62. https://doi.org/10.1097/ACM.0000000000003704.
30. Hanley K, Gillespie C, Zabar S, Adams J, Kalet A. Monitoring communication skills progress of medical students: establishing a baseline has value, predicting the future is difficult. Patient Educ Couns. 2019;102(2):309–15. https://doi.org/10.1016/j.pec.2018.09.010. Epub 2018 Sep 13
31. Chicoine J, Nallamadi S, Kudlowitz D, Sherpa N, Summar A, Noyes M, Crotty K. Examining effects of an accelerated preclerkship curricula on communication skill development in medical students. In: Society of general internal medicine annual meeting poster presentation, April 2024, Boston, MA; 2024.

32. Cangiarella J, Eliasz K, Kalet A, Cohen E, Abramson S, Gillespie C. A preliminary evaluation of students' learning and performance outcomes in an accelerated 3-year MD pathway program. J Grad Med Educ. 2022;14(1):99–107. https://doi.org/10.4300/JGME-D-21-00284.1. PMID: 35222827; PMCID: PMC8848877.
33. Frades I, Matthiesen R. Overview on techniques in cluster analysis. Methods Mol Biol. 2010;593:81–107. https://doi.org/10.1007/978-1-60327-194-3_5.
34. Becker-Haimes EM, Lushin V, Creed TA, Beidas RS. Characterizing the heterogeneity of clinician practice use in community mental health using latent profile analysis. BMC Psychiatry. 2019;19(1):257. https://doi.org/10.1186/s12888-019-2234-0. PMID: 31443697; PMCID: PMC6708227.
35. Altshuler L, Wilhite JA, Hardowar K, Crowe R, Hanley K, Kalet A, Zabar S, Gillespie C, Ark T. Understanding medical student paths to communication skills expertise using latent profile analysis. Med Teach. 2023;45(10):1140–7. https://doi.org/10.1080/0142159X.2023.2193303. Epub 2023 Mar 24
36. Cronbach LJ, Nageswari R, Gleser GC. Theory of generalizability: a liberalization of reliability theory. Br J Stat Psychol. 1963;16:137–63.
37. Blood AD, Park YS, Lukas RV, Brorson JR. Neurology objective structured clinical examination reliability using generalizability theory. Neurology. 2015;85(18):1623–9. https://doi.org/10.1212/WNL.0000000000002053. Epub 2015 Oct 2. PMID: 26432851; PMCID: PMC4642144.
38. Martin IG, Jolly B. Predictive validity and estimated cut score of an objective structured clinical examination (OSCE) used as an assessment of clinical skills at the end of the first clinical year. Med Educ. 2002;36(5):418–25. https://doi.org/10.1046/j.1365-2923.2002.01207.x.
39. Pugh D, Bhanji F, Cole G, Dupre J, Hatala R, Humphrey-Murto S, Touchie C, Wood TJ. Do OSCE progress test scores predict performance in a national high-stakes examination? Med Educ. 2016;2016(50):351–8. https://doi.org/10.1111/medu.12942.
40. Wallenstein J, Heron S, Santen S, Shayne P, Ander D. A core competency-based objective structured clinical examination (OSCE) can predict future resident performance. Acad Emerg Med. 2010;17(Suppl 2):S67–71. https://doi.org/10.1111/j.1553-2712.2010.00894.x.
41. Soong TK, Ho CM. Artificial intelligence in medical OSCEs: reflections and future developments. Adv Med Educ Pract. 2021;12:167–73. https://doi.org/10.2147/AMEP.S287926. PMID: 33628074; PMCID: PMC7899303.
42. Jani KH, Jones KA, Jones GW, Amiel J, Barron B, Elhadad N. Machine learning to extract communication and history-taking skills in OSCE transcripts. Med Educ. 2020;54(12):1159–70. https://doi.org/10.1111/medu.14347. Epub 2020 Sep 10

Scholarship and Education Research Registry

7

Lisa Altshuler, Jeffrey Wilhite, Colleen Gillespie, Adina Kalet, and Sondra Zabar

Highly trained standardized patients can provide a rich source of data for health professions education research. There has been substantial exploration of the methodology itself and its use in studying the development of learner skills and documenting impact of educational initiatives and curricula [1–3]. In this chapter, we provide a brief overview of how to facilitate research with standardized patients and describe a range of research questions and methods which may be relevant to your educational efforts and setting.

Overview of Medical Education Research

OSCEs are a valuable research tool to study research questions about individual learners and programmatic impact, as well as other questions relevant to medical education. OSCEs can be used as an outcome measure, or existing OSCE data gathered as part of learners' educational experience may be used to answer research questions about the process of education. Medical education research (MER) calls for investigators to develop a theory, obtain or prepare data, and secure ethnical approval prior to analyses and dissemination (Fig. 7.1).

L. Altshuler (✉) · J. Wilhite · S. Zabar
Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA
e-mail: lisa.altshuler@nyulangone.org

C. Gillespie
Institute for Innovations in Medical Education, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA

A. Kalet
Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_7

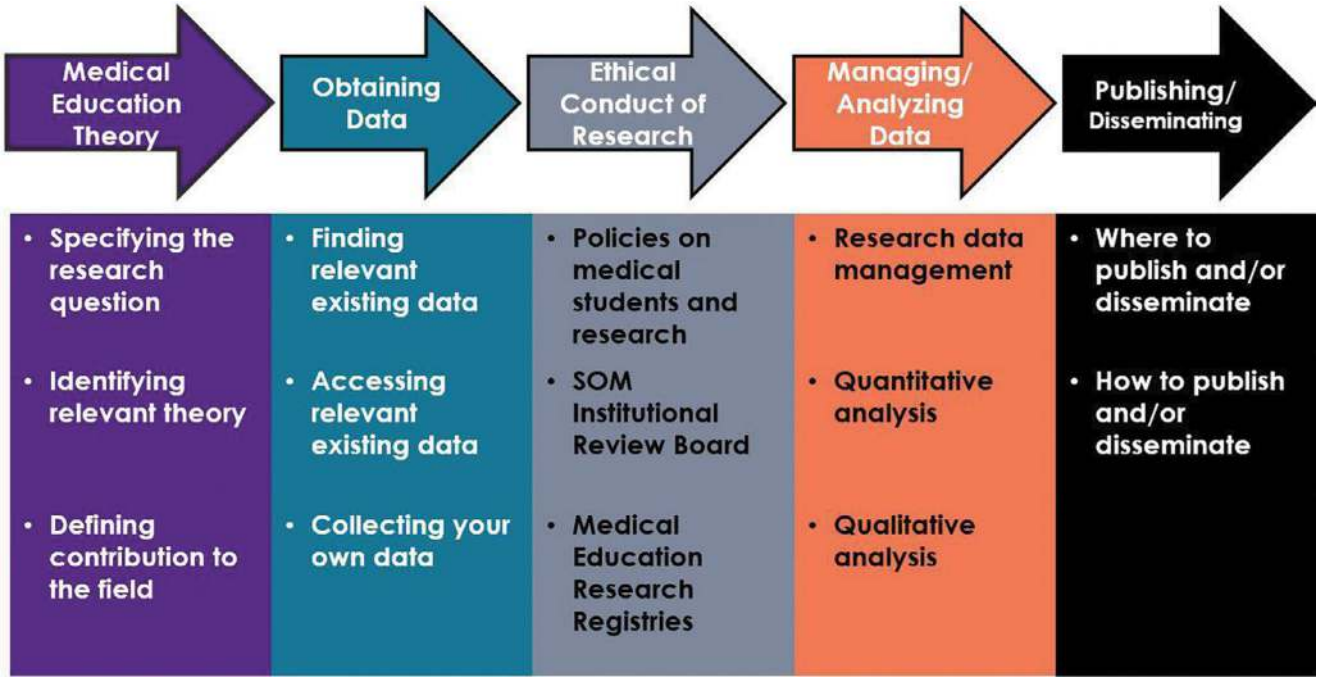


Fig. 7.1 Medical education research

Implementation and Facilitation of Medical Education Research (MER)

Conducting meaningful MER can be challenging. Sample sizes are often too small to address complex research questions, and it is difficult to gather consistent and comparable data gathering across either different training programs or across the UME and GME continuum. Additionally, gaining approval for human subjects research requires navigating the institutional review board (IRB) in medical environments that are more schooled in biomedical than educational research issues. Implementation of a MER registry can address these challenges and enhance the research process in multifaceted ways [4–7]. This is an IRB approved registry that allows for use of data routinely collected for educational purposes. At its core, a MER registry can serve as a centralized repository of meticulously curated, linkable, ethically approved research data. Researchers can access research-ready datasets, larger than those available to them directly, without having to negotiate the often-cumbersome administrative procedures associated with gathering complex data.

Development and implementation of a MER registry requires building collaborations among medical educators, administrative leadership, trainees, and education researchers. Consensus must be developed about policies regarding data-sharing, data linkage, consent processes, measurement approaches, dissemination of findings, and author credit, among other issues. For example, we had (and have ongoing) discussions to develop shared approaches to the measurement of communication and other clinical skills, forming the structure and consistency of rating checklists. We have worked toward ensuring policies about reliable and consistent scoring and all other aspects of measurement validity [8]. Additionally, the MER registry is more than just a repository; it functions as a learning collaborative for researchers navigating the complex landscape of medical education inquiry. By providing insights and direction for the formulation of research questions, research methods, and data analysis, the registry offers invaluable support to investigators. Encouraging the utilization of frameworks such as PICOT (see below) further enhances the clarity and focus of research inquiries, fostering more structured and productive research.

Implementation of a MER registry requires establishing a robust consent mechanism, typically integrated into new trainee orientation events to inform trainees about the registry and obtain written consent. This ensures that data collection adheres to rigorous ethical standards while laying the groundwork for high-quality educational research and institution-wide professional development and career advancement. The success of a MER registry hinges on the dedication of skilled personnel tasked with management and oversight of both the data gathering and use processes. A core contact person, or data steward, plays a pivotal role in expediting dataset preparation. This streamlined approach not only accelerates the analysis and reporting processes but also builds institutional capacity for educational scholarship and mitigates duplicative efforts among researchers.

The establishment of a MER registry marks a significant milestone in the advancement of medical education research. Through meticulous attention to consent mechanisms, dedicated personnel, streamlined data preparation processes, and guidance for research inquiries, institutions can cultivate a collaborative and research-driven environment that fosters innovation and excellence in medical education practice.

Crafting OSCE-Related Research Questions: Using PICOT Format

In the sections below, we describe two general categories of research questions focusing on OSCE performance. The first addresses the use of the OSCE itself as a measurement tool assessing learner performance. As with any measurement approach, there are strengths and drawbacks in using OSCEs. An assessment of the benefits of using such a methodology depends on the context and the content of the performance addressed in each instance, along with the specific learners assessed. Measurement quality issues such as standardization, reliability, and validity are critical. These are discussed below in part I. The second broad category of research questions utilizes OSCEs to answer specific educational questions. This might include efforts to assess learners, document the development and epidemiology of skills over time, and examine the impact of educational efforts or specific curricula. Part II provides an overview of such questions as well as examples from our own body of research. The range of research questions is only limited by your curiosity and creativity.

To describe the research questions, we utilize the PICOT format [9]. This is an evidence-based framework that is widely used in clinical and medical education research. It helps design research questions that are specific, focused, and answerable and facilitates the search for related evidence and the appraisal of research findings. PICOT identifies the essential components of a well-developed research question as follows:

P (Population)—Define the specific target population of learners to be studied.

I (Intervention or independent variable)—Clearly describe the educational strategy or educational variable (e.g., specific skill or target studied) to be evaluated and its potential impact on the target population.

C (Comparison)—Identify any comparison group, which allows researchers to draw meaningful conclusions about intervention efficacy.

O (Outcome)—Specify the outcome measures to be gathered, ensuring that they are clearly defined and aligned with study objectives.

T (Time Frame)—Specify the time frame for the study to allow adequately for the evaluation of the duration of effects over time.

See Table 7.1 for sample PICOT worksheet for your own research ideas.

I. OSCEs as a Measurement Tool

Research can focus on aspects of the OSCE as a measurement tool or on the efficacy and validity of the OSCE in addressing learners' clinical skills and competence. Ensuring that the assessment is reliable and valid, particularly if it is a high-stakes assessment for the trainee or program, and is essential (Item A below).

Using best practices for implementing OSCEs and understanding and documenting the strengths and potential weaknesses of OSCEs in different contexts before conducting the research guides the research design and provides cautions for the interpretation of the results (Items B–D). Considering these items also helps educators distinguish between formative OSCEs, which might not require rigorous measurement quality, and summative experiences, which are more consequential for trainees and therefore require attention to ensure high-quality measurement.

Table 7.1 PICOT worksheet for use when designing an SP-based research question

Study title:		
Research question:		
Population	***Target group/sample of learners***	
Intervention/ Independent variable	***"Treatment" (curricular intervention, innovation, or variable expected to affect outcome)***	
Comparison	***Reference group (counterfactual –what happens without intervention?)***	
Outcome	***Measure reflecting effectiveness of intervention***	
Time	***Duration for study; when outcome is measured***	

(A) **Validation Studies**

Establishing the validity of a measure requires an accumulation and organization of evidence to support the argument that the measure is appropriate for use as intended or "fit for purpose" (e.g., to make promotion decisions) (see Cook et al. [10], Kinnear et al. [11], and Ark et al. [8] for well-established frameworks to guide validity testing). Reliability is the essential bedrock for a validity argument. Reliability studies might include (1) identifying rigorous protocols for ensuring uniformity of the OSCE scenarios (including consistency of case details, SP portrayal, and implementation details like timing, learner instructions, and other logistics), (2) establishing inter-rater agreement between or among raters observing (i.e., standardized patients, faculty observers) the same clinical encounter and/or (3) intra-rater reliability agreement between ratings made by the same rater on the same performance (e.g., one in-person and the second on a recording), (4) consistent use of a measurement instrument for a rater or group of raters using the measurement across a group of trainees, and (5) establishing that the measurement instrument consistently measures the construct as it was designed to do (e.g., as a scale, measured by Cronbach's alpha). Once a measure has an acceptable level of reliability or consistency, many types of validity studies are possible. For instance, OSCE results can be compared with other established measures of clinical competence, such as faculty ratings on clinical rotations, observations of real-world performance, or other validated assessment tools, or a range of outcomes of the OSCE can be identified and described.

Example A study assessing the ability of an OSCE to predict future trainee performance in actual clinical settings by comparing the scores of medical students on the OSCE with their subsequent clinical performance in real patient settings.

Example An investigation—both quantitatively and qualitatively—into the inter-rater agreement among standard patients observing the same set of clinical encounters.

(B) **Feedback Studies**

Feedback is a crucial component of the learning process and development of expertise. If properly designed, OSCEs are a very high-value opportunity to study the impact of feedback on all stakeholders including individual learners, health professions education program leaders, and even SPs.

Example Studies of the impact of different feedback delivery methods in OSCEs on subsequent performance and skill retention should aim to explore various aspects of the feedback, including timing (e.g., immediate or delayed), format (e.g., written, verbal, or both), and content (e.g., specificity, areas for improvement, and guidance on how to enhance performance). Additional questions might explore learner's and educator's perceptions of feedback, the impact of types of feedback on learners' subsequent performance, the effect of feedback from OSCEs on the curriculum, or the potential of OSCE participation influencing SP's personal growth and health behaviors [12].

(C) **Implementation Studies**

These studies investigate the practical aspects of implementing OSCEs in medical education, considering logistical challenges, cost-effectiveness, and the overall feasibility of integrating OSCEs into curricula.

Example A study examining the logistical challenges, benefits, and trade-offs of integrating OSCEs into a medical school curriculum, considering factors such as resources, faculty training, and student feedback.

(D) **Qualitative Studies**

Some research designs involve qualitative methods, such as interviews or focus group discussions, to explore learners' experiences with OSCEs, perceptions of the assessment process, and the impact of OSCE feedback on their learning.

Example Holding a focus group to explore medical students' opinions on the OSCE experience, including their attitudes toward standardized participants, the authenticity of scenarios, and the impact of feedback.

Creating Scholarship from Your OSCE Program

In our institution, OSCEs have served as a launching point for new measures of trainee competence, including entrustment ([13], telemedicine-specific practices [14], and specific communication challenges (e.g., vaccine hesitancy [15], implicit bias [16], responding to social determinants of health [5–7]. As the second column of Table 7.2 details, during refinement of measures for our simulated "Night on-Call" program (see Chap. 10 and Kalet et al. [17] for a full description of the UME near-graduate comprehensive, immersive simulation), a single item assessing the degree to which an OSCE station observer "trusts" their trainee to engage in independent practice was piloted [13]. Pilot results suggested that entrustment scores provide nuance to other measures of preparation for the transition to residency. This entrustment data, when collected longitudinally, can provide ongoing assessment of the overall trustworthiness of residents—a construct that residency program directors prioritize

Table 7.2 PICOT examples of validity studies

Study:	**Ob/Gyn Test of Integrated Professional Skills** [18]	**Night on-Call: Readiness for Residency Simulation** [13]
Research question	**How do assessments of residents' procedural and communication skills in an OSCE correspond to faculty ratings?**	**Can multi-source assessment of OSCE performance be used to make "entrustability" judgments?**
Population	Ob/Gyn residents	Fourth year medical students
Intervention/ Independent variable	N/A	N/A
Comparison	Faculty ratings	Underlying "construct"—Derived from multiple trainee assessments
Outcome	Validity of OSCE measure	Reliable and valid "entrustability" decision regarding readiness for residency
Time	Cross-sectional	Cross-sectional

Table 7.3 PICOT examples of needs assessments

Study:	**Virtual Communication Skills** [19]	**Virtual Urgent Care Onboarding** [20]
Research question	**Do primary care residents perform differently during an OSCE case conducted in-person vs. virtually?**	**Do physicians know how to conduct virtual triage and communicate with a patient with upper respiratory infection?**
Population	Medicine residents (PGY 1–3)	Virtual urgent care physicians as part of a faculty onboarding program
Intervention/ Independent variable	Telemedicine curriculum	Description of telemedicine Communication skills
Comparison	Residents with and without exposure to virtual care practice opportunities	N/A
Outcome	Comprehensive video visit with patient, appropriate communication skill use	Appropriate communication skill use, patient education, physical exam skills, medication reconciliation
Time	1 year between assessment points	Cross-sectional

when discussing the UME-GME handoff. This measure has since been added to every annual OSCE at our institution. The second column of table 7.2, "Ob/Gyn Test of Integrated Professional Skills," describes validated measures of trainee competence by comparing to faculty ratings.

II. OSCEs to Explore Patterns of Development, Guide Educational Efforts and Ensure Clinical Quality

OSCE data can elucidate the development of skills over time (e.g., typical patterns of communication skills development from beginning to end of medical school), assessing the impact of curriculum as well as curricular gaps, identifying weaknesses in individual learners, and ensuring quality assurance.

(E) Needs Assessment

Needs assessments help investigators identify areas that their learners would benefit from additional training opportunities in. From a study design standpoint, they are usually descriptive, but there are opportunities for quasi-experimental study efforts based on the structure of your research question. For example, examining data across multiple years is a quasi-experimental design that looks at learner over time or group change.

Example Table 7.3: "Virtual Communication Skills" exhibits the quasi-experimental example of analyzing data collected during two annual assessments, one in-person and the other conducted virtually [14, 19]. Exploration of performance during the two periods contributed to curricular refinement efforts.

Example Examples of a needs assessments include our baseline explorations of resident willingness to engage in gender affirming care [21], virtual care provision and telemedicine skill integration among urgent care physicians ([20]; Table 7.3: "Virtual Urgent Care Onboarding"), assessment of resident ability to identify impaired colleagues [22], and whether residents can identify and manage opioid overdose [23]. Through these approaches, one can either identify areas of individual strength and weakness or identify in the aggregate needs for training or remediation in your institution.

(F) Educational Epidemiology

Rigorous studies of learners over key components of the continuum are relatively new to medical education.

Longitudinal studies track the performance of learners over an extended period, using OSCEs at multiple points to assess the development and retention of clinical skills over time.

Within our institution, we have amassed over 15 years of OSCE data. Because the communication skills checklist we have used was designed rigorously, used ubiquitously, and stably maintained over time, we are now able to analyze the data for trends.

Example In Table 7.4: "Communication Skills over Time," we identified cohorts of trainees in need of remediation (a rare event) and have ascertained the specific measures that are most meaningful in providing insight on our trainees' skills [24] and have learned that refining skills is difficult [25]. For example, the checklist items "Allowed you (the patient) to talk without interrupting" and "nonverbal behavior enriched communication" differentiate between average and low-performing profiles of learners, while patient education items are challenging across all learner profiles [24]. We have also compared practice patterns between OSCE stations and USP visits and found patient education skills demonstrated when a trainee is a student aren't always transferred to independent practice as residents [5–7].

Example A longitudinal study tracking the development of clinical skills in a cohort of medical students over several years, using OSCEs at different points to measure progress. **Cohort studies**, examining groups of learners over time, use OSCEs to provide insights into trends, patterns, and factors influencing clinical competence development. They can also assess the impact of "nested" interventions such as changes in curriculum.

Example A cohort study following a group of students during medical school and residency, utilizing OSCEs to identify trends in competency development and analyzing factors influencing progress ([26]; Table 7.4: "Communication Skills Across the UME-GME Continuum").

(G) **Curriculum Evaluation**

Researchers may use OSCEs to evaluate the impact of specific interventions on learners' clinical skills and performance.

Example An intervention study implementing a targeted training program for communication skills and assessing its impact. We have been able to evaluate the impact of curriculum on home visit skills [27], pain management ([28]; Table 7.5: "Pain Assessment and Management Skills After Exposure to Experiential Curriculum"), substance use disorder treatment [29], and transgender health skills ([30]; Table 7.5: "Impact of OSCEs Across the UME-GME Continuum on Transgender Health Skills") using our standardized checklist. Using OSCEs early in an education program (we do a three-case OSCE the second week of medical school) has the distinct advantage of establishing a baseline for skills. A pre-intervention OSCE is invaluable for studying change in response to curricular activities; however, the researcher needs to recognize that the OSCE itself may have raised a student's awareness to what is expected of them and therefore influenced post-intervention performance independent of the intervention. This "testing effect" can be detected by adding a control—"non-intervention" arm to the study design.

(H) **Quality Assurance**

SP data can also provide new insight on care quality. Here, we define care quality as the standards for clinical triage, management, and follow up. Through ongoing OSCE assessments, we are able review learner performance on a number of quality measures, including prescribing patterns and referrals for services. As an example, our OSCE stations on counselling asthmatic and opioid-dependent patients gave clear insight into clinical practices of our residents. Additionally, our use of unannounced SPs (USPs) in tandem with OSCEs provides insight into our trainees' clinical practice patterns

Table 7.4 PICOT examples of longitudinal studies

Study:	**Communication Skills Over Time** [24, 25]	**Communication Skills Across the UME-GME Continuum** [26]
Research question	**How do learners' communication skills develop over time? Are there distinct trajectory patterns?**	**How do competencies develop between medical school and residency?**
Population	Medical students (into residency)	Students who complete both medical school and residency at NYU
Intervention/ Independent variable	Training/skills development	Our training programs (medical school and/or residencies); patterns of competence development
Comparison	Variations in education and training (cohort effects); learner characteristics	Performance during OSCEs at two developmental time points (medical student post-clerkship assessment vs. PGY2 assessment)
Outcome	Communication skills	Trainee skills during PGY2
Time	Longitudinal—Repeated measures at three points during medical school (developmental)	Longitudinal (cohort type study)

Table 7.5 PICOT examples of curriculum evaluation

Study:	**Impact of OSCEs across the UME-GME Continuum on Transgender Health Skills** [30]	**Pain Assessment and Management Skills After Exposure to Experiential Curriculum** [28]
Research question	**Does prior exposure to transgender health skills during medical school improve competency with these skills during residency?**	**Does exposure to experiential pain management curriculum improve?**
Population	Internal medicine residents	Third year medical students
Intervention/ Independent variable	Transgender OSCE during medical school	Pain management curriculum
Comparison	Residents who trained at NYUGSOM and who previously participated in a transgender OSCE as part of their UME curriculum vs. residents who attended different medical schools and did not participate in NYUGSOM transgender OSCE	Medical students exposed to curriculum vs. prior year's class which did not receive it
Outcome	Skills demonstrated on OSCE case	Skills demonstrated on OSCE case
Time	Single-year	Cross-sectional

Table 7.6 PICOT examples of skills transfer for quality assurance

Study:	**Opioid Prescribing Practices** [31]	**Variation in Depression Management** [32]
Research question	**Are residents following best practices in opioid prescribing?**	**How much variation is there in prescribing, ordering, referrals, documentation, and scheduling of return visits when managing depression?**
Population	Medicine residents (USP visits)	Medicine residents (USP visits)
Intervention/ independent variable	Assessment of alignment with opioid prescribing recommendations	Description of variations in care
Comparison	Exploration of performance based on specialty, PGY level, clinical experience	Degree of variation; patterns of variation (by resident); influences on variation (case; setting; hecticness, etc.)
Outcome	USP measures of assessment, education and counseling, and treatment recommendations	Chart review of prescriptions, orders, referrals, documentation, return visit scheduling for same clinical cases
Time	Cross-sectional; monitor over time	Repeated measures (multiple cases over time)

[5–7]. In one such study, our USP data captured asthma medication prescribing, resident patient education practices, medical record documentation, and appropriate time for return to clinic for a patient presenting for acute exacerbation [5–7]. Through similarly structured research questions, we examined variation in residents' prescribing practices when patient requested opioids for back pain ([31]; Table 7.6: "Opioid Prescribing Practices"). We also examined variation in resident diagnosis, treatment, and prescribing practices for USPs presenting with depression symptoms ([32]; Table 7.6: "Variation in Depression Management"). In your own work, care quality-based questions will help you incorporate more complex, nondescriptive questions into your hypotheses. In these experimental studies, SP data can help you understand if, how much, and under what circumstances skills transfer from OSCEs into practice.

III. Other Research Questions: Hypothesis Testing and Causal Modeling (Example: Do Trainee or SP Behaviors Impact the Outcome?)

We used OSCE data to quantify the relationship between patient education skills and degree of trust for a vaccine-hesitant SP [15]. Hypothesis testing and causal modeling will ultimately allow you to identify specific skills needed for effective care provision, and results can inform you about an individual learner's needs. Other examples in our institution include our study of attitudes toward obesity and their impact on communication skills ([33]; Table 7.7: "Role of Physician Attitudes Toward Obesity") and the impact of unique smoking cessation communication styles on patient activation ([34]; Table 7.7: "Impact of Communication Style on Activation"). Predicting specified outcomes, such as asso-

Table 7.7 PICOT examples of hypothesis testing and causal modeling

Study:	**Role of Physician Attitudes Toward Obesity** [33]	**Impact of Communication Style on Activation** [34]
Research question	**Is competence in assessing and counseling obese patients about weight management associated with attitudes toward obesity?**	**How do communication styles impact resident ability to sway a patient who smokes?**
Population	Practicing physicians	Residents
Intervention/ independent variable	Attitudes toward obesity (obese patients, causes of obesity, treatment expectations)	Quality of patient education efforts
Comparison	Physicians with positive vs. negative attitudes	Patient education coupled with direct cessation recommendation vs. solely recommending cessation
Outcome	Obesity counseling practices (assessment, counseling, management)	Degree of patient activation
Time	Cross-sectional	Cross-sectional

ciations between skills and trust, between clinical chaos and communication, or between care provision and clinical outcome, may be the next frontier for medical education research.

Conclusions

In summary, OSCEs can be used as needs assessments in medical education and healthcare settings to identify gaps in clinical knowledge and skills, assess competency levels, tailor educational interventions, evaluate training program effectiveness, inform curriculum development, and ensure quality assurance. Medical education researchers should seek to share these insights focusing on problem-solving or discovery of insights that build on current knowledge and are generalizable. Data generated by SPs through highly reliable and consistent use of checklists has the potential to impact educational programs and add to our understanding of clinical competence. Though data are often collected purely for educational purposes, with careful theory and evidence informed pre-planning, the data can also be purposed for meaningful, publishable scholarship. While not an absolute requirement, a MER registry can optimize the research processes across your institution. Use of the PICOT framework will greatly enhance team alignment and output and bolster your research enterprise.

Best Practices

- Create a shared mental model for assessment across learner types and programs to facilitate educational scholarship and research.
- Invest in infrastructure to make sure that, when appropriate, the same measurements are used over time and across programs.
- Use the PICOT format to help create questions and design medical education studies to foster successful scholarship.
- Collaborate with your Office of Medical Education and IRB to establish a medical educational research registry to create a thriving community of practice for educational scholarship.

References

1. Chang O, Holbrook AM, Lohit S, Lohit S, Deng J, Xu J, Lee M, Cheng A. Comparability of Objective Structured Clinical Examinations (OSCEs) and written tests for assessing medical school students' competencies: a scoping review. Eval Health Prof. 2023;46(3):213–24. https://doi.org/10.1177/01632787231165797.
2. Kreptul D, Thomas RE. Family medicine resident OSCEs: a systematic review. Educ Prim Care. 2016;27(6):471–7. https://doi.org/10.1080/14739879.2016.1205835.
3. Montgomery A, Chang HR, Ho MH, Smerdely P, Traynor V. The use and effect of OSCES in post-registration nurses: an integrative review. Nurse Educ Today. 2021;100:104845. https://doi.org/10.1016/j.nedt.2021.104845.
4. Gillespie C, Zabar S, Altshuler L, et al. The Research on Medical Education Outcomes (ROMEO) registry: addressing ethical and practical challenges of using "bigger" longitudinal educational data. Acad Med. 2016;91(5):690–5. https://doi.org/10.1097/ACM.0000000000000920.
5. Wilhite JA, Altshuler L, Zabar S, Gillespie C, Kalet A. Development and maintenance of a medical education research registry. BMC Med Educ. 2020a;20(1):199. Published 2020 Jun 19. https://doi.org/10.1186/s12909-020-02113-5.
6. Wilhite JA, Fisher H, Altshuler L, Cannel E, Hardowar K, Hanley K, Gillespie C, Zabar S. Gasping for air: measuring patient education and activation skillsets in two clinical assessment contexts. BMJ Simul Technol Enhanc Learn. 2020b;7(5):428–30. Published 2020 Nov 27. https://doi.org/10.1136/bmjstel-2020-000759.
7. Wilhite J, Hardowar K, Fisher H, Porter B, Wallach A, Altshuler L, Hanley K, Zabar S, Gillespie C. Clinical problem solving and social determinants of health: a descriptive study using unannounced standardized patients to directly observe how resident physicians

respond to social determinants of health [published correction appears in Diagnosis (Berl). 2020 Oct 27;8(1):129. doi: 10.1515/dx-2020–0125]. Diagnosis (Berl). 2020c;7(3):313–24. https://doi.org/10.1515/dx-2020-0002.
8. Ark T, Kalet A, Tewksbury L, et al. Validity evidence for the clinical communication skills assessment tool (CCSAT) from 9 years of implementation in a high stakes medical student OSCE. Patient Educ Couns. Published online June 2; 2024. https://doi.org/10.1016/j.pec.2024.108323.
9. Riva JJ, Malik KM, Burnie SJ, Endicott AR, Busse JW. What is your research question? An introduction to the PICOT format for clinicians. J Can Chiropr Assoc. 2012;56(3):167–71. PMID: 22997465; PMCID: PMC3430448
10. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. Med Educ. 2015;49(6):560–75. https://doi.org/10.1111/medu.12678. PMID: 25989405
11. Kinnear B, Kelleher M, May B, Sall D, Schauer DP, Schumacher DJ, Warm EJ. Constructing a validity map for a workplace-based assessment system: cross-walking Messick and Kane. Acad Med. 2021;96(7S):S64–9. https://doi.org/10.1097/ACM.0000000000004112
12. Plaksin J, Nicholson J, Kundrod S, Zabar S, Kalet A, Altshuler L. The benefits and risks of being a standardized patient: a narrative review of the literature. Patient. 2016;9(1):15–25. https://doi.org/10.1007/s40271-015-0127-y.
13. Eliasz KL, Ark TK, Nick MW, Ng GM, Zabar S, Kalet AL. Capturing entrustment: using an end-of-training simulated workplace to assess the entrustment of near-graduating medical students from multiple perspectives. Medical Science Educator. 2018;28(4):739–47. https://doi.org/10.1007/s40670-018-0628-0.
14. Sartori DJ, Hayes RW, Horlick M, Adams JG, Zabar SR. The TeleHealth OSCE: preparing trainees to use telemedicine as a tool for transitions of care. J Grad Med Educ. 2020;12(6):764–8. https://doi.org/10.4300/JGME-D-20-00039.1.
15. Wilhite JA, Zabar S, Gillespie C, et al. "I Don't Trust it": use of a routine OSCE to identify core communication skills required for counseling a vaccine-hesitant patient. J Gen Intern Med. 2022;37(9):2330–4. https://doi.org/10.1007/s11606-022-07495-4.
16. Gonzalez CM, Ark TK, Fisher MR, Marantz PR, Burgess DJ, Milan F, Samuel MT, Lypson ML, Rodriguez CJ, Kalet AL. Racial implicit bias and communication among physicians in a simulated environment. JAMA Netw Open. 2024;7(3):e242181. Published 2024 Mar 4. https://doi.org/10.1001/jamanetworkopen.2024.2181.
17. Kalet A, Zabar S, Szyld D, Yavner SD, Song H, Nick MW, Ng G, Pusic MV, Denicola C, Blum C, Eliasz KL, Nicholson J, Riles TS. A simulated "Night-onCall" to assess and address the readiness-for-internship of transitioning medical students. Adv Simul (Lond). 2017;14(2):13. https://doi.org/10.1186/s41077-017-0046-1.
18. Winkel AF, Gillespie C, Hiruma MT, Goepfert AR, Zabar S, Szyld D. Test of integrated professional skills: objective structured clinical examination/simulation hybrid assessment of obstetrics-gynecology residents' skill integration. J Grad Med Educ. 2014;6(1):117–22. https://doi.org/10.4300/JGME-D-13-00055.1. PMID: 24701321; PMCID: PMC3963767
19. Boardman D, Wilhite JA, Adams J, et al. Telemedicine training in the COVID era: revamping a routine OSCE to prepare medicine residents for virtual care. J Med Educat Curri Develop. 2021;8:23821205211024076. Published 2021 Jun 16. https://doi.org/10.1177/23821205211024076
20. Sartori DJ, Lakdawala V, Levitt HB, Sherwin JA, Testa PA, Zabar SR. Standardizing quality of virtual urgent care: using standardized patients in a unique experiential onboarding program. MedEdPORTAL. 2022;18:11244. https://doi.org/10.15766/mep_2374-8265.11244
21. Greene RE, Hanley K, Cook TE, Gillespie C, Zabar S. Meeting the primary care needs of transgender patients through simulation. J Grad Med Educ. 2017;9(3):380–1. https://doi.org/10.4300/JGME-D-16-00770.1. https://doi.org/10.15766/mep_2374-8265.11244.
22. Zabar S, Hanley K, Horlick M, et al. "I cannot take this any more!": preparing interns to identify and help a struggling colleague. J Gen Intern Med. 2019;34(5):773–7. https://doi.org/10.1007/s11606-019-04886-y.
23. Lynn M, Calvo-Friedman A, Hanley K, Wilhite J. Can residents identify and manage opioid overdose? Med Educ. 2020;54(11):1060–1. https://doi.org/10.1111/medu.14340.
24. Altshuler L, Wilhite J, Hardowar K, Crowe R, Hanley K, Kalet AL, Zabar S, Gillespie C, Ark T. Understanding medical student paths to communication skills expertise using latent profile analysis. Med Teach. 2023;45(10):1140–7. https://doi.org/10.1080/0142159X.2023.2193303.
25. Hanley K, Gillespie C, Zabar S, Adams J, Kalet A. Monitoring communication skills progress of medical students: establishing a baseline has value, predicting the future is difficult. Patient Educ Couns. 2019;102(2):309–15. https://doi.org/10.1016/j.pec.2018.09.010.
26. Gillespie C, Beltran CP, Nallamaddi S, Hardowar K, Wilhite JA, Hanley K, Crowe R, Horlick M, Porter B, Kalet A, Zabar S, Ark TA, Altshuler L. Communication skills across the UME-GME continuum: Information Gathering and Patient Education and Counseling Improve from Medical School to Residency. In Abstracts from the 2024 Annual Meeting of the Society of General Internal Medicine. J Gen Intern Med. 2024;39(S2)
27. Zabar S, Hanley K, Adams J, Ark TK. No place like home: teaching home visits. Med Educ. 2009;43(11):1102–3. https://doi.org/10.1111/j.1365-2923.2009.03487.x.
28. Stevens DL, King D, Laponis R, et al. Medical students retain pain assessment and management skills long after an experiential curriculum: a controlled study. Pain. 2009;145(3):319–24. https://doi.org/10.1016/j.pain.2009.06.030.
29. Reich H, Hanley K, Altshuler L. Block of addiction medicine (BAM!): an intensive resident curriculum improves comfort with substance use disorders. J Gen Intern Med. 2019; Conference: S790–S791
30. Beltran CP, Wilhite JA, Hayes RW, LoSchiavo C, Crotty K, Adams J, Hauck K, Crowe R, Kudlowitz D, Katz K, Gillespie C, Zabar S, Green R. Practice makes perfect: objective structured clinical examinations across the UME-to-GME continuum improve care of transgender simulated patients. J Grad Med Educ. 2024;16(2):182–94. https://doi.org/10.4300/JGME-D-23-00573.1
31. Hanley K, Zabar S, Altshuler L, et al. Opioid vs nonopioid prescribers: variations in care for a standardized acute back pain case. Subst Abus. 2017;38(3):324–9. https://doi.org/10.1080/08897077.2017.1319894.
32. Zabar S, Hanley K, Watsula-Morley A, et al. Using unannounced standardized patients to explore variation in care for patients with depression. J Grad Med Educ. 2018;10(3):285–91. https://doi.org/10.4300/JGME-D-17-00736.1.
33. Jay M, Kalet A, Ark T, et al. Physicians' attitudes about obesity and their associations with competency and specialty: a cross-sectional study. BMC Health Serv Res. 2009;9:106. Published 2009 Jun 24. https://doi.org/10.1186/1472-6963-9-106.
34. Wilhite JA, Velcani F, Watsula-Morley A, et al. Igniting activation: using unannounced standardized patients to measure patient activation in smoking cessation. Addict Behav Rep. 2019;9:100179Published 2019 Mar 28. https://doi.org/10.1016/j.abrep.2019.100179.

Performance-Based Assessment in Practice

8

Sondra Zabar, Kathleen Hanley, Harriet Fisher, Renee Heller, Jennifer Adams, Adina Kalet, and Colleen Gillespie

Performance-based assessments (PBAs) are an ideal methodology of training and assessment that contributes to a multitude of individual and programmatic benefits. They can provide the opportunity to practice solving common pragmatic dilemmas as well as expose learners to uncommon clinical scenarios. PBAs also facilitate program evaluation by identifying programmatic strengths and gaps and allow educators to address core curricular deficiencies. OSCE trainings and assessments can be nimbly adapted as treatment protocols emerge or change, while annual and longitudinal OSCE programs can produce useful data for administrative dashboards and scholarship. All health professions educators can draw on their creativity and clinical experience to create cases that enrich the learning environment and ensure competency, and the process of creating cases and assessment forms and engaging faculty observers for feedback builds both a shared mental model for assessment and a community for learning.

OSCEs are well-suited to address myriad educational goals. In this chapter, we showcase instances where OSCEs can (1) teach and assess new curricular topics, (2) ensure everyone has exposure to essential skills, (3) provide a program of assessment for trainees, and (4) prepare clinicians for transitions. Within these four categories, Table 8.1 gives an overview of programs that use OSCEs in each of these four contexts. These OSCEs are further described in Tables 8.2, 8.3, 8.4 and 8.5, and bolded cases have been "spotlighted" to demonstrate a diverse set of programs with reflections from educators on their process, the impact of their program, their advice on implementation, and lessons learned. Tables 8.2, 8.3, 8.4 and 8.5 offer a curated catalog of approaches or projects with literature references for further exploration ("Selected References" section at the end). We hope this chapter will spark your creativity and lead to innovations to help prepare the next generation of health professionals.

S. Zabar (✉) · K. Hanley · H. Fisher · R. Heller · J. Adams
Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA
e-mail: sondra.zabar@nyulangone.org

A. Kalet
Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA

C. Gillespie
Institute for Innovations in Medical Education, Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_8

Table 8.1 Road map for curated catalog of OSCEs—organized by type and spotlighted programs

1. OSCEs to teach new curricular topics	3. OSCEs as programs of assessment—monitoring your clinicians' skills development
• **Virtual heath at the veterans association ambulatory clinic** • Addressing discriminatory patient comments • Social determinants of health screening in ambulatory care • Palliative care fellowship • Emergency medicine faculty: Starting buprenorphine in the emergency department • Telemedicine Care at Student Health Center • Assessing COVID-19 preparedness	• **Neurology residency** • **Gastroenterology fellowship** • Primary care internal medicine residency program • Internal medicine residency program • Pediatric residency • Surgery residency
2. OSCEs to teach/assess skills "everyone" needs	**4. OSCEs to prepare learners for transitions**
• **Care for Transgender and Gender Diverse (TGD) people** • **OBGYN residency** • Reproductive endocrinology fellowship • Research assistants recruiting older adults to clinical trials • Transitions of care management skills	• **Onboarding new faculty** • Anesthesiology residency program • Neonatal intensive care fellows • First night-onCall: Experiential patient safety orientation for all incoming interns • Night-onCall: Near-graduation assessment of all 13 EPAs prior to internship

Section 1: OSCEs to Teach New Curricular Topics

Spotlight on OSCEs for Virtual Heath at the Veterans Association Ambulatory Clinic

Interview with Anne Dembitzer, MD, with the New York Harbor Veterans Health Affairs and Zoe Phillips, Research Coordinator

Why Did You Choose to Implement This Program in the Clinical Setting During Patient Care?

In the 2 and a half years since the onset of the pandemic, many of our residents and faculty at the Veterans Health Affairs Clinic (and most other hospital systems) have expressed varying levels of comfort with aspects of virtual care (e.g., relationship building with patients, communication over video), and it has become clear that quality telehealth care requires a set of distinct skills. Because both residents and faculty have limited time to complete trainings, we designed a workplace-based training that would occur during clinic time and provide direct, actionable feedback on their ability to provide telehealth care. Our learners were primary care residents and faculty at the VA, so many of their patients were older adults with chronic conditions that could be well-served and monitored via virtual care. Identifying hearing loss and adjusting sound and screens were particularly important skills for clinicians to connect with VA patients. We designed cases that would allow our learners to practice interactions with real patients and referral to VA-specific virtual care resources and services.

How Was This Announced Standardized Patient (ASP) Experiential Learning Program Received by Learners and Clinical Leadership?

The announced standardized patient (ASP) experiential learning program was very useful for clinicians and provided insight into clinicians' skill development; 92% of clinicians reported it was well designed and engaging, and 88% agreed it was an effective way to reinforce good habits in healthcare communication. Though some were frustrated that the encounter took place during clinic time, 69% said they would recommend it to a peer. Learners who performed well on telemedicine skills were more likely to be considered good communicators, which supported our assertion that telemedicine skills were a critical part of quality care. In the future, we might want to collect ongoing data to determine whether feedback had a positive impact on clinicians' telehealth skills. We only sent one visit to each clinician, so we didn't get the opportunity to measure change. We also had no way of providing feedback on the EMR note the clinician was asked to complete. In the future, we could use a standardized checklist to assess the EHR note and give feedback to the clinicians.

What Advice Do You Have for Someone Looking to Create a Similar ASP Program?

(1) *Find out what faculty want and need.* Conduct a needs assessment with target learners and educators to develop cases.
(2) *Promote efficiency of ASPs*. Inform the clinical community about the advantages of work-placed learning—using the EHR and team in real time.
(3) *Find a clinical partner*. Schedule announced standardized patient visits at regular intervals with clinicians and have a clinical partner who understands the flow of the clinic.

(4) *Don't forget the notes*. Develop a method for evaluating clinical notes to provide immediate feedback to clinicians.

(5) *Feedback, feedback, and more feedback*. Ensure feedback to clinicians is timely (within a few days) of the encounter so the clinician can recall the reason for the feedback.

Table 8.2 OSCEs to teach new curricular topics

Program/topic	Problem	Logistics	Cases	Outcomes
Virtual Heath at the Veterans Association Ambulatory Clinic	• The transition to telemedicine at the onset of the COVID-19 pandemic revealed an urgent need for virtual health skills • Implementing a competency-based experiential training program using Announced Standardized Patients (ASPs) in the workplace with the actual Electronic Health Record (EHR) could rapidly train residents and faculty	**Learner**: 76 VA providers (16 faculty and 60 residents) **Frequency:** one visit/learner, over 4 months **Feedback**: SP verbal feedback and checklist, learner completed a note and self-refection **Location**: in-situ: ASP placed in providers regular clinic schedule, visit took place on clinic's video platform	**Number of cases**: 2 (1 per learner) **Core Topic**: Chronic disease management of hypertension with either a: – Hearing-impaired patient – Patient with low technology skills	• Workplace-based simulation acts as a needs assessment and quality improvement program to teach a new set of skills • Embedded into regular clinic as a virtual visit for a new patient, modeling real life experience • Feedback went to clinical and educational leadership as well as learners • Collaborative project with IT, EHR, Residency program and GIM faculty • Signaled importance of these skills since the training took place during dedicated clinical time
Addressing Discriminatory Patient Comments	• Patients demanding for reassignment based on biases and discrimination is a fairly common occurrence that residents are unprepared to address • Lack of team support after experiencing discrimination compounds the initial insult, further affecting the individual's well-being and ability to effectively deliver patient care	**Learner**: Pediatric faculty and residents, IM Residents, Neurology residents **Frequency:** once during residency **Feedback**: SP verbal feedback and checklist, faculty observer feedback **Location**: Simulation center, in-person or virtual	**Number of cases**: 1 **Core Topic:** Case centered on supervising a Muslim Standardized Learner (SL) who experienced discrimination after an assigned patient requested a new physician. The SL rated participants on three domains: – Supervision – Relationship Development – Supporting the Learner	• This OSCE signaled leadership's support of learners and that they took the institution's policies seriously • Case can be performed as a station in an OSCE, part of workshop, or as a group OSCE • Cases can be adapted for faculty, residents or medical students • A qualitative analysis of comments from the SL revealed that high-performing residents (HPR) more frequently demonstrated positive behaviors (using supportive body language, verbalizing support) and elicited feelings of validation • All learners noted that the opportunity to practice these skills in a safe environment was valuable given the charged subject matter

(continued)

Table 8.2 (continued)

Program/topic	Problem	Logistics	Cases	Outcomes
Social Determinants of Health (SDOH) Screening in Ambulatory Care	• Little is known about how clinical care teams respond to SDOH issues during a primary care visit and how best to increase the response to a SDOH once identified • Note: used Unannounced Standardized Patients (USPs) = secret shoppers	**Learner**: 2 primary care clinics received 108 visits **Frequency:** 4 visits scheduled/week, for 18 months **Feedback**: SP checklist, individual reports to learner and aggregate reports to clinical leadership **Location**: in-situ, Primary Care Clinic	**Number of cases**: 6 different cases **Core Topics** Address financial and housing SDOH embedded in 5 clinical cases: – Wellness Visit – Fatigue – Asthma exacerbation – Patient Education and Hepatitis B – Acute Back Pain	• Cycles of audit and feedback with targeted educational information reinforced best practices • Rates of identifying housing insecurity increased for teams that received audit/feedback reports. • Partnership with residency directors and clinical leadership (across disciplines) was crucial to the success of a USP program • Project led to new collaborations with Chief Medical Office and additional funding.
Palliative Care Fellowship	Extubating is a common but not predictable event that requires competence by anesthesiology, primary care, geriatric and emergency medicine physicians	**Learner**: 28 total: PC, Pulmonary and Geriatric fellows, Neurosurgery residents, and PC NPs **Frequency:** Annual **Feedback**: SP verbal feedback and checklist, faculty observer feedback, all formative **Location**: Simulation center, in-person	**Number of cases**: 1 **Core Topic**: Hybrid OSCE and simulation: SP, ventilator, and mannequin on palliative extubating	• Funded by internal program for medical education innovation grants, then integrated and sustained in curriculum for all 3 fellowships • Won Best Education Innovation at the 2020 American Thoracic Society meeting • Created SIM-plifying Palliative Extubating Toolkit for other programs to be able to implement the simulation
Emergency Medicine (EM) Faculty: Starting Buprenorphine in the Emergency Department (ED)	• Initiating buprenorphine and referral in the ED for treatment for Opioid Use Disorder (OUD) is feasible, highly effective, and cost-effective treatment, yet is rarely initiated in EDs • EM faculty have a range of comfort and experience in prescribing treatment for OUD • Important, timely topic for faculty development	**Learners:** EM Faculty **Frequency**: Part of an annual faculty development series **Feedback:** SP checklist, faculty observer feedback, and 20 min group debrief with faculty facilitator after each GOSCE case **Location**: Simulation center, in-person	**Number of cases**: 3 Group OSCE (GOSCE) cases **Core Topics**: • Patient anxious to leave the ED after treatment with intranasal naloxone by EMS • Patient requesting detox admission from "Oxy" • Patient requesting/ demanding opioid pain medication In each case they: a) Assess for ED-initiated buprenorphine b) Discuss the patients' substance use c) Provide counseling and education where appropriate	• The 3 cases provided intense simulated experiences with challenges often faced in the ED • The structured debrief focused on enhancing faculty's ability to properly screen and treat patients with buprenorphine as well as improving communication skills • The program increased faculty- reported comfort administering buprenorphine, helped facilitate the use of non-stigmatizing language, gave faculty new strategies when confronted with OUD in the ED, and positively changed perceptions of buprenorphine and OUD

(continued)

Table 8.2 (continued)

Program/topic	Problem	Logistics	Cases	Outcomes
Telemedicine Care at Student Health Center	• As practices shift toward telemedicine, many lack clinical workflow for virtual care provision • Private practices do not have routine opportunities for continuing medical education for their staff • Secret shopper/ Unannounced Standardized Patient (USP) visits can be used for quality improvement projects	**Learner**: 16 clinicians (8 MDs, 7NPs, 1 PA) **Frequency:** 2 cases/ learner **Feedback**: USP checklist; chart review, individual reports, aggregate reports for clinical leadership **Location**: in-situ; USP placed in provider's regular clinic schedule, used clinic's video visit platform	**Number of cases**: 2 **Core Topics**: • Sexual health screening and counseling • Depression recognition and management	• This simulation program created an academic partnership with clinical faculty that served as a quality improvement project for a private practice/student center • Provided data and feedback to clinical leadership and individual clinicians • Highlighted gaps in workflow, especially compared to in-person visits
Assessing COVID-19 Preparedness	• The emergence of COVID-19 necessitated rapidly identifying and isolating potentially infected individuals • USPs presented a unique way to evaluate the level of COVID-19 preparedness in a community clinic	**Learner:** 4 community health clinical teams **Frequency:** 1 visit/ team; 4 total visits over 1 month **Feedback**: Immediate feedback from USP to team when USP discloses identity after team isolates them; USP checklist **Location**: in-situ, Ambulatory Clinic	**Number of cases:** 1 **Core Topic:** USP was trained to present at each team's front desk with the complaint of feeling unwell; the USP then evaluated each step of care throughout the encounter	• Case was developed rapidly with clinical leadership to evaluate whether teams were appropriately following protocols for isolation • USPs revealed significant variation in care practices within a clinical system • The immediate feedback allowed the teams to take quick steps to ensure all team members understood the protocol and could provide the most appropriate care to a potentially infected individual

Section 2: OSCEs to Teach/Assess Skills "Everyone" Needs

Spotlight on Care for Transgender and Gender Diverse (TGD) People

Interview with Richard Greene, MD, MHPE, Associate Director of the Primary Care Internal Medicine Residency Program at NYUGSOM and Director of Bellevue Hospital LGBTQ+ Health Services.

What Motivated You to Use OSCEs to Teach and Assess Caring for Trans* Patients?

When I began teaching learners about the skills of gender-affirming care, it quickly became clear that an important barrier between the provider and the patient was discomfort. Providers are often uncomfortable because they lack experience taking care of transgender or gender diverse (TGD) patients. This can lead them to feel uncomfortable with everything from their language to their ability to answer questions. Similarly, given the health disparities among TGD patients and the fact that the vast majority of TGD patients have had negative experiences when accessing healthcare, it is important not to throw inexperienced learners in to see patients who may have high needs and difficult prior experiences. An OSCE is the perfect way to fill that gap—learners have the experience of practicing language and receiving feedback from TGD standardized patients (SPs) and develop some confidence, and the *patients* get better care when they come to their clinical spaces.

What Have You Learned?

We have learned A LOT doing these OSCEs. The most important lesson we learned is that good psychosocial skills are the same for everyone but don't always cover the necessary material. Learners with strong psychosocial skills received high marks from the SPs even when they side stepped any specific questions about their gender identity (when these were directly relevant to the care the patient would receive). We had to develop more specific items to assess whether the learners were discussing gender identity-related topics with their patients in ways that made the SPs feel they were getting the best care possible. We also learned that we need two types of cases with TGD SPs, one set to test

the skills and knowledge of gender-affirming care and another to observe how learners interact with TGD SPs when the clinical aspects of the case do not involve gender identity. We don't want learners to ignore gender identity when it is relevant but don't want them to focus exclusively on a patient's gender identity to the exclusion of the actual medical issue.

What Advice Do You Have for Someone Looking to Create a Similar OSCE Program?

Given the vulnerability for harm to TGD people by the healthcare system, there are a few things we recommend always:

(1) *Be clear what skills you want learners to demonstrate, and be sure the case will take them there and the checklist reflects that.* We have cases to demonstrate how to have a risk/benefit conversation about starting gender-affirming hormones and some to demonstrate how to apologize when someone is misgendered. These are both critical but very different skills!

(2) *Always engage TGD people to portray these cases.* Using cisgender people to portray transgender patients can miss critical information from the person's lived experience. The content of the case will be different from the person who is portraying it but holding a shared identity will add critical authenticity.

(3) *Remember that portraying cases that involve your identity can be emotionally draining!* Check in with SPs and give frequent breaks to re-center. One SP noted, "It's really hard when you see that look in their eye that says, 'I don't know how to take care of you' in the moment you disclose your identity in the case, when that's my real identity, too."

(4) *As you train community members to participate in these OSCEs, set expectations with them about the level of the learners.* Also allow them to discuss prior trauma or frustrations they may have experienced in healthcare settings so they don't unintentionally project that onto learners or into their feedback and that they feel emotionally ready to participate in the case repeatedly.

(5) *Include TGD people in all aspects of an OSCE project.* Have a review process for cases that includes TGD people and include TGD people on any education research and publications to ensure the voice of TGD people is reflected in messages that go out to the medical community.

(6) *Remember that this may be your learners' first encounter with someone who identifies as TGD, and if the encounter is too difficult or too embarrassing for them, it may dissuade them from wanting to care for TGD people in the future*. One of the goals of our cases is always to increase the learner self-efficacy!

Table 8.3 OSCEs to teach/assess skills "everyone" needs

Program/topic	Problem	Logistics	Cases	Outcomes
Care for Transgender and Gender Diverse (TGD) people	• Curriculum gap for care of transgender patient in UME, GME, and CME • Not possible to ensure every trainee will get the experience of caring for TGD people during training due to learning environment or number of learners • Necessary to protect TGD patients from negative or exploitive care experiences with learners who do not yet have the necessary skills for affirming care	**Learner**: First year medical students, IM, EM, and GYN residents, pulmonary and GI fellows **Frequency:** Annual **Feedback**: SP verbal feedback and checklist, faculty observer feedback **Location**: Simulation center, in-person or virtual * **Notes**: Transgender individual hired and trained to portray case	**Number of cases**: 10 **Core topics:** Cases adapted for specialty and learner stage— • Trans-inclusive language (affirmed name, pronouns, etc.) • Clinical knowledge (TGD-inclusive screening and medical needs) • Initiating hormone therapy, continuation of hormone therapy, managing adverse effects • Abdominal pain with organ inventory • Taking an inclusive and affirming sexual history • Managing sexual health	• Developed a group of "patients as teachers"—TGD-identified individuals to portray the cases authentically • Simulation implemented for potential low-frequency clinical experience ensured everyone has training and competency in DEI • Can be used in remote format for expanded dissemination • Residents reported they were grateful for the opportunity to practice the skills. Faculty and SP feedback helped residents identify learning needs, including how to ask directly about transition, appropriate terminology, how to admit lack of knowledge to a patient, and putting aside preconceptions about a patient's identity

(continued)

Table 8.3 (continued)

Program/topic	Problem	Logistics	Cases	Outcomes
OBGYN Residency	• Procedural residents must be competent not only with technical skills but the discussions that occur with patients • No standardized approach to assess both procedures and communication challenges • Opportunity to design cases that use task trainers and SP in each case	**Learner**: 20 residents/year **Frequency:** Annual **Feedback**: SP checklist, faculty observer feedback individual summary reports, annual residency programmatic evaluation **Location**: Simulation center, in-person	**Number of cases**: 5/year per PGY **Core topics:** Each case has a communication or professional challenge and procedure skill: • Ultrasound/delivering bad news • Suturing/vaginal laceration • Operative delivery/informed consent • Evidence-based medicine/clinical question • Genetic testing/family meeting	• Standard faculty evaluation did not correlate with OSCE scores; OSCE scores showed much more variation • Debriefing with faculty provided an important interface for identification of performance gaps and individualization of learning plans • Initial funding was provided by internal grants, but the program was able to be sustained as an annual educational activity for PGY1 and PGY3 even after funding period ended • The annual simulation assessment program was integrated into a simulation director role and identified as adding value for faculty development and scholarship
Reproductive Endocrinology Fellowship	• Fellows faced with many high-stress, difficult conversations yet no explicit common curriculum or standardized method for feedback. • These types of conversation are everyday occurrences in their practice	**Learner**: 6 fellows and advanced practice professionals **Frequency:** Annual **Feedback**: SP verbal feedback and checklist, faculty observer feedback For 2 out of 3 cases, individual and program summary reports **Location**: Simulation center, in-person	**Number of cases**: 3 **Core topics:** • Low egg discovery • Breaking news of a miscarriage • Complex logistics and emotional patient	• Used case template and educational consultation with simulation center to create faculty-identified cases • Serves as model for local program and other health professionals (NPs and PAs) • Collaborated with regional program to increase numbers • Learners recognized that these difficult conversations are advanced skills but will become common, everyday tools for their careers • Abstract on simulation accepted for plenary presentation
Research assistants recruiting older adults to clinical trials	• There are common core skills for recruitment of participants in clinical trials • Research staff have limited opportunity for standardized experiential learning. • Older adults may have unique reservations that should be addressed during recruitment to clinical trials • Recruitment to clinical trials is essential for the institution's research	**Learner**: 60 research staff at academic medical center, approx. 9–12/session **Frequency:** 3 times/year **Feedback**: SP verbal feedback, faculty trainer feedback, group structure enables peer feedback and discussion **Location**: Simulation center, in-person or virtual	**Number of cases**: 3 **Core topics:** • Identifying and supporting a hearing-impaired patient. • Addressing family members' role in decision-making. • Recognizing and acknowledging previous experiences and distrust of research	• Provided research staff with a realistic, engaging, and reproducible opportunity to practice recruitment • 92% reported the activity taught them something new; 98% reported it provided valuable feedback; 100% said they would like to participate again • Provided research coordinators with information on strengths and weaknesses of team members skills • Cases and debriefing can be adapted to align with patient population and/or specifics of clinical trial • Funding provided by: NIH/NIA for the NYU engagement in longevity and medicine (ELM) research collaborative (1R24AG063725) • Potential to sustain program through institution's research core

(continued)

Table 8.3 (continued)

Program/topic	Problem	Logistics	Cases	Outcomes
Transitions of care management (TCM) skills	• The post-hospital discharge period is a vulnerable time as patients have increased risk of clinical deterioration, readmission, and mortality • Residents should be able to conduct effective post-discharge visits to avoid preventable issues • We need to know TCM baseline skills and curriculum gaps as this is a relatively new rotation	**Learner**: 43 residents across 3 PGY years **Frequency:** One visit/learner during the spring clinic block **Feedback**: SP verbal feedback and checklist, individual and program summary reports **Location**: In situ: Announced standardized patient (ASP) placed in providers regular transitions of care clinic rotation schedule, visit took place on clinic's video platform	**Number of cases**: 1 **Core topics:** • Telemedicine skills. • Transitions of care management: – Making appropriate referrals – Asking about symptoms and medication since discharge • Case-specific skills: – Recognizing medication error on discharge • Making appropriate plan for lifestyle modification	• Demonstrated that ASP encounters are an effective tool for practicing TCM and telemedicine skills and providing feedback to residents • Resident's baseline could improve for case-specific, TCM, and telemedicine skills, indicating potential areas for preceptors to focus on • These visits provided a needs assessment to inform curriculum updates, which can be assessed by repeating the intervention to measure improvement • This workplace-based assessment was successfully integrated into regular clinical practice • The EHR and clinical documentation note was also analyzed, revealing the need for further education to improve telemedicine transition visits • Awarded foundation grant to further develop, evaluate, and disseminate project

Section 3: OSCEs as Programs of Assessment—Monitoring Your Clinicians' Skills Development

Spotlight on Neurology Residency

Interview with Arielle Kurzweil, MD, Director of the Neurology Residency Program at NYUGSOM

How Did You Choose to Structure/Develop Your OSCE?

When I first became program director of the NYU Neurology Residency, I reflected on my time in medical training and thought about what was missing. When I was a medical student, we had courses and workshops focused on humanistic aspects of medicine and communication. But once I was a resident, I focused so much on neurologic disease, diagnosis, and management and rarely thought about how to build communication skills.

As a program director, I really began to appreciate that interpersonal communication and professionalism are ACGME core competencies! I was being asked to determine whether my residents were competent in these domains but had no real standardized way to do this. OSCEs are a phenomenal way to observe and assess trainee communication skills in real time and provide immediate specific feedback for their professional growth. Furthermore, simulation guarantees a universal experience among all trainees, in a safe environment.

A committee of faculty educators in our department thought about difficult conversations we have encountered in our careers and created a survey for our residents to assess which of these topics they had encountered, were observed doing, and received feedback on. Based on these results, we created an OSCE curriculum for all years in residency training that involves four OSCEs per year and increases in complexity and relevance of topic.

What Are Some Key Takeaways from Implementing Your OSCE Program?

Our OSCE curriculum has taught us a great deal about "scripts." And not scripts in the sense of writing the OSCE but rather a set of phrases, vocabulary, and ideas that can be taught and developed and then applied to real-life scenarios to make care more effective. Our curriculum involves having difficult conversations with patients, their family members, and colleagues. Some examples include providing a terminal neurologic diagnosis to a patient, discussing a poor prognosis with a family member after a

patient suffers hypoxic injury to the brain, and giving feedback to a colleague you are working with. We have noticed that our trainees have applied the "scripts" they develop in the OSCEs to other similar scenarios, and this has created more comfort surrounding difficult or uncomfortable conversations.

Our OSCE curriculum has also created a terrific opportunity for both trainees and faculty to engage more with scholarly work in the education research realm. The overall curriculum as well as the individual OSCEs have been presented as posters and platforms presentations at institutional, regional, and national meetings. Many of them have been published. We have even adapted our OSCE and implement as faculty development for continuing medical education credit.

What Advice Do You Have for Someone Looking to Create a Similar OSCE Program?

First and foremost, we highly recommend creating a set of OSCEs for each year of residency training that build on earlier skills and increase in complexity and nuance. This can lead to a comprehensive program of performance-based assessment. A few pearls that we have learned along the way:

(1) *Find your gaps!* Start with a needs assessment to inquire about skills that your trainees want to develop and do not feel they get the opportunity to practice under observation.
(2) *OSCE is a team sport.* Get as many different faculty involved as possible to write cases and instruct learners. It leads to more engagement for education endeavors in the department and creates more opportunity for scholarship.
(3) *Plan.* Book the simulation center/rooms well in advance!
(4) *Maximize your time.* Consider running each set of OSCEs over different days in two separate block rotations to accommodate those on night float and vacation; call and ensure all your resident learners can participate.
(5) *Be flexible!* If times are changing and new skills seem necessary to focus on, adjust OSCEs accordingly. For example, the COVID-19 pandemic shed light on the need to adjust to a virtual environment for taking care of patients and engaging with colleagues. Therefore, since 2020, we have made at least one OSCE station in each year of the curriculum as a remote case, to allow trainees to practice these skills.
(6) *Listen to your learners.* After each set of OSCEs, we survey our learners about whether they found the educational exercise helpful, if there are any ways to improve it, and what other topics might be useful to explore using the OSCE format in the future. These surveys have led to great programmatic improvements, and new OSCE cases have been developed based on trainee feedback.

Spotlight on Gastroenterology Fellowship

Interview with Elizabeth Weinshel, MD, and Renee Williams, MD, MHPE, former Directors of the Gastroenterology Fellowship Program at NYUGSOM

How Did You Choose to Structure/Develop Your OSCE?

Our motivation for starting a program with standardized patients (SPs) was to expose the GI fellows to high-intensity, high-impact, and/or low-frequency patient care situations such as apologizing for a complication, breaking bad news, and caring for patients with cultural differences from their own. These programs provide a "safe space" for the fellows to practice challenging interactions and receive constructive feedback from the SPs and from some of their key clinical faculty. This provides information to the program directors about the fellows' performances in what are often infrequent but important scenarios. Our faculty used a faculty feedback guide similar to the SP checklist which enables them to be more specific in their feedback. This was especially useful for our junior faculty. We invited GI training programs in NYC to send their fellows (and faculty who served as observers) creating a community of practice for both our fellows and educators. We were able to use the results from the SPs (not from faculty, who are not standardized or specifically trained) to give the training program directors in our region additional information about the fellows' performances to use during their semiannual feedback sessions.

How Has Your OSCE Program Changed Over Time?

Our program has persisted and grown over the past decade to include other regional GI training programs as well as other subspecialties. We have worked with pediatric gastroenterologists, endocrinologists, and pulmonary and critical care physicians to create programs for their learners. We also developed a program for faculty (unobserved except for the SPs) which includes reviewing videos of their own performance for faculty development and CME/MOC credits. We've added new scenarios allowing us to add subspecialty specific cases (i.e., inflammatory bowel disease and transplant hepatology.)

We shared all our experience and created a comprehensive manual that resides on the American College of Gastroenterology's medical education portal for use by GI training programs nationwide. Faculty have leveraged their experiences with us to launch new educational programs that have elevated their career paths and provided leadership opportunities in their home institutions.

In the COVID era, we utilized Zoom to eliminate travel time and to enable participants from programs further from our home site to become. Lastly, we identified aspects from each case that align with the ACGME milestones (specifically in the domains of patient care, interpersonal communication, professionalism, and systems-based practice) which allowed us to develop "report cards" with scores that correlate to the milestones. The program directors found this to be helpful during their ACGME accreditation survey.

What Advice or Best Practices Do You Have for Someone Looking to Create a Similar OSCE Program?

(1) *Partner with others.* It was incredibly helpful for us to align with another group who pioneered these programs at our institution. We learned from them and then shared materials and standardized patients so that others could replicate the program.

(2) *Use exiting assessment frameworks.* We were fortunate to have validated checklists for the various communication and professionalism domains available for our use. We have found that more often than not, those doing similar work in other specialties or institutions are willing to share materials and expertise.

(3) *OSCEs can be done in many different ways and locations.* Having a state-of-the-art simulation center was also incredibly helpful, although not required.

Table 8.4 OSCEs as programs of assessment—monitoring your clinicians' skills development

Program/topic	Problem	Logistics	Cases	Outcomes
Neurology Residency	• No standardized assessment of communication skills or professionalism • Lack of shared mental model for program of performance-based assessment of residents and curriculum	**Learner**: 60 residents (16–20/PGY) **Frequency**: Annual **Feedback**: SP verbal feedback and checklist, faculty observer feedback **Location**: Simulation center, in-person	**Number of cases**: 4 cases/PGY (12 cases/year) **Core topics:** • Informed consent (medication for acute stroke) • Difficult colleague • Struggling trainee • Teleneurology • Disclosing medical error • Giving feedback • Functional neurologic disorder • Uncomfortable patient encounter • Delivering bad news • Telestroke	• Adapted cases from GME case bank for time efficiency • Using the same cases each year means that the first year is the most intense work and the program becomes routine in subsequent years • Faculty lead for each PGY and case generated collaborative leadership • Created scholarship opportunities to increase abstract submissions and papers • Enabled longitudinal performance data on individual and curriculum • Established program as a national leader in resident education with focus on simulation
Gastroenterology Fellowship	• Small number of fellows to create large-scale simulation • No national performance-based assessment curriculum for GI fellows on communication and professionalism skills	**Learner**: 30 residents (8–12/PGY) combined with intra-city program involving 3 to 4 programs **Frequency:** Annual **Feedback**: SP checklist, faculty observer feedback Individual summary reports **Location**: Simulation center, in-person or virtual	**Number of cases**: 4 cases/PGY (12 cases/year) **Core topics:** All with focus on communication skills: • Inflammatory bowel disease • Systems-based practice • Liver disease • Pediatric professionalism	• Lead faculty developed expertise that was recognized locally and nationally • Employed a systematic approach to writing each case which involved a medical student or resident with the goal of creating an abstract/publication • Led to collaboration with GI society for publication of program director guide of cases and national CME course • Elevated medical education reputation in GI division
Primary care Residency	• Limited curriculum evaluation besides learner satisfaction. • No standardized performance assessment or opportunity for progressive testing • Infrequent direct observation of clinical work	**Learner**: 30 residents (10/PGY) **Frequency:** Annual event over 1–2 days **Feedback**: SP checklist, faculty observer feedback **Location**: Simulation center, in-person or virtual	**Number of cases:** 7–10 cases/year; 4 new cases/year written by faculty **Core topics:** • Teaching • Behavioral change • Behavioral health • Telemedicine • Clinical reasoning	• Provides annual program evaluation and progressive testing • Enabled individual reports for clinical competency committee • Cases shared with other residency programs (e.g., informed consent, interprofessional communication) • Created scholarly opportunities • Used as an opportunity to add just-in-time topics during the pandemic such as vaccine hesitancy, COVID-19, and telemedicine

(continued)

Table 8.4 (continued)

Program/topic	Problem	Logistics	Cases	Outcomes
Internal medicine Residency	• Difficult to standardize curriculum in large program • PGY1s needed clinical experience in common situations that they are uncommonly observed doing, PGY2s needed observation in less common, high-stakes situations • Lack of standardized performance data for clinical competency committee (CCC) meetings	**Learner**: 60 residents (30 PGY1s and 30 PGY2s) **Frequency:** Annual event for PGY1 and PGY2 during ambulatory block **Feedback**: SP checklist, faculty observer feedback **Location**: Simulation center, in-person or virtual	**Number of cases**: 5 cases/year (usually the same each year for PGY1 and PGY2s) **Core topics:** • Shared decision-making • Advanced directives • Struggling colleague • Management of common conditions • Disclosure of medical errors	• Created shared mental model for faculty on feedback and core clinical skills • Ensured that there is a common curriculum across a large program • Identified learners who need additional instruction/coaching in first year and provided follow-up in second year • Allowed for video review when needed for remediation • Used for clinical competency committee meetings
Pediatric Residency	• Need for a longitudinal curriculum which addresses parents' perspective • Difficult to get parent or patient perspective in the clinical setting • Grant catalyzed the opportunity to develop a telemedicine/ communication curriculum	**Learner**: 50 residents/year (over 4–6 sessions) **Frequency:** 4 cases/year **Feedback**: SP checklist, faculty observer, and family faculty observer feedback **Location**: Simulation center, in-person or virtual	**Number of cases**: 4 **Core topics**: Developed 2 cases/ year with the goal of deploying different cases each year to establish a case bank over time Bank includes: • Medical error • Bad news delivery • Intentional harm • Failure to thrive	• Facilitated partnership with Institute for Family and Child Centered Care which contributes to SPs funding • Integrated communication curriculum with family faculty (parents who are trained to give feedback) • Provides an annual, 360-degree evaluation of residents • Published on checklist and simulation to build national reputation • Curriculum incorporated as standard part of residency even after post-grant funding
Surgery Residency	• Challenge of teaching and assessing the professionalism and communication competencies for the Accreditation Council for Graduate Medical Education (ACGME), especially for surgical residencies • Need to evaluate the effectiveness of the new faculty-designed surgical professionalism in clinical education (SPICE) seminar series in order to demonstrate assimilation of skills and improved performance	**Learner**: 15 surgical residents **Frequency:** 6 station OSCE completed pre- and post-SPICE curriculum **Feedback**: SP checklist **Location**: Simulation center, in-person	**Number of cases**: 6 • Hernia repair began on the wrong side—Surgical mistake • Delivering bad news • Use of interpreter during surgical informed consent • Identifying a struggling colleague • Transfer call for patient from ER to surgery • Health proxy	• Provided evidence that professionalism and communication competencies can be taught • Base cases adapted for other residencies (EM, IM, OBGYN) • Reinforced that communication skills are as important as surgical technical skills—After the longitudinal SPICE curriculum, professionalism skills and confidence in recommending residents to other patients significantly improved • Used for ACGME accreditation • Presented nationally and became model for other surgical programs

Section 4: OSCEs to Prepare Learners for Transitions

Spotlight on OSCE for Onboarding New Faculty

Interview with Andrew Wallach, MD, Ambulatory Care Chief Medical Officer at NYC Health + Hospitals and Associate Director of Clinical Innovations and Clinical Affairs for the Division of General Internal Medicine at NYUGSOM, Katherine Hochman, MD, Director of the Division of Hospital Medicine at NYUGSOM, and Renee Heller, Program Coordinator.

Why Did You Create an OSCE for New Faculty?

Newly recruited clinicians have heterogeneous clinical training and experiences; therefore, it is important to evaluate competencies and establish expectations for patient communication, safety, and performance. We have administered an experiential simulation program as part of the onboarding requirements for new faculty since 2017. Recent emerging pathogen outbreaks presented a unique opportunity to explore just-in-time education for clinicians during a public health crisis. Although initially designed to reinforce communication skills standards, we now also see our onboarding simulation as an opportunity to train new internists and hospitalists on appropriate counseling of patients about emerging public health issues. We have recently leveraged this program to disseminate knowledge about the opioid epidemic, COVID-19, and Mpox to all new faculty.

What Has Been the Impact of Your Program?

To date, over 200 faculty and advanced practice providers have participated in our onboarding program. The program

creates a shared mental model of the role of our division in faculty education and ongoing learning, introduces key divisional contacts and resources, and brings together hospitalist and ambulatory clinicians from different sites. Given that each of our site directors have made this training mandatory, it signals the importance our division places on communication skills and ongoing education. Faculty remember this event years later!

The results of our post-program feedback surveys from trainings involving a COVID-19 or Mpox case show that this experiential onboarding program is valuable to newly hired clinicians and effectively informs a large number of clinicians during emerging public health crises. Ninety-five percent of participants who completed the program agreed that the program helped them feel more confident about counseling a patient about emerging public health issues. Case-specific education (like COVID-19 or Mpox knowledge and experience) and patient activation may be connected. Our findings suggest that targeted simulation education could be key to preparing clinicians during unfamiliar disease outbreaks and ensuring that patients receive high-quality care.

What Advice Do You Have for Someone Looking to Create a Similar Onboarding Program?

(1) *Build a team.* Work collaboratively with clinical and educational leaders to create a program which will help facilitate integration into your health system's particular workflow. Identifying clinical champions at each care delivery site is essential.

(2) *Provide SP feedback as close as possible to the event.* Create simple mechanisms to deliver feedback within a week of program to be most effective. It is important that this feedback is not viewed by the learner's direct supervisor in order to create a safe learning space.

(3) *Make the program annual.* The heavy lift will mostly occur in the first year, and eventually, the program will become part of the institutional culture.

(4) *Contribute to larger conversations at your institution and nationally.* Align new cases with public health issues, institutional initiatives, or quality improvement needs to garner support and recognition. This increases the positive reception of both clinical leaders and of the learners, as they will value the opportunity to improve their clinical care in especially relevant ways.

Table 8.5 OSCEs to prepare learners for transitions

Program/topic	Problem	Logistics	Cases	Outcomes
Onboarding new faculty	• New faculty come with heterogeneous clinical training and experiences from medical school, residency training, and prior jobs • Onboarding is an opportunity for new faculty to understand their new institution's expectations and standards for patient communication, safety, and performance • Public health emergencies such as COVID-19 and Mpox involve diseases with symptoms, treatment, and protocols that rapidly evolve and are often unfamiliar to clinicians	**Learner**: New faculty (< 18 months) Offered to inpatient and outpatient faculty; MD'S and APP's **Frequency:** Annual, new clinicians are assigned **Feedback:** SP verbal feedback and individual checklist report, group debrief with faculty facilitator **Location**: Simulation center or virtual	**Number of cases**: 3 **Core topics:** • Correcting a medical error • Conducting a post-discharge follow-up call and counseling when a patient brings up an unexpected concern (e.g., COVID-19, Mpox) • Giving feedback to a struggling resident	• Opportunity to introduce new faculty to each other, both within and across sites, as well as meet divisional leadership and administrative staff • Group debrief modeled lifelong learning, provided a safe learning environment, and an opportunity to reflect on one's own practice and learning goals • Faculty receive an institutional resource guide containing relevant resources and contacts, as well as the SP checklist data to reflect on their performance and make a commitment to change • Ability to adapt the onboarding program to address future public health crises when just-in-time physician education is especially important

(continued)

Table 8.5 (continued)

Program/topic	Problem	Logistics	Cases	Outcomes
Anesthesiology residency Program	• New national performance-based assessment for licensing was implemented in 2016 • Residencies were introducing OSCEs as practice tests to prepare for licensing	**Learner**: 21 residents **Frequency:** Annual **Feedback**: SP verbal feedback and checklist, faculty feedback, individual summary reports **Location**: Simulation center, in-person	**Number of cases**: 4 **Core topics:** • Discuss an ethical issue • Obtain informed consent • Address periprocedural complications • Teach problem-based learning	• Used national exam blueprint to design cases with simulation center medical educator • Learners reported feeling prepared for national exam and appreciating direct observation and feedback • Program leadership found summary reports helpful and now will start implementing OSCEs for more than just licensing exam preparation
Neonatal Intensive care Fellows	• Difficult conversations are a common task. National organization encourages programs to use simulation as a teaching tool • Fellows come with different skills from a diverse set of residency programs. • Neonatal National Society's case bank did not have assessment forms	**Learner**: 6 fellows **Frequency** annual, at the start of fellowship **Feedback**: SP verbal feedback and checklist, faculty observer feedback **Location**: Simulation center, in-person	**Number of cases**: 2 **Core topics:** Adapted two cases from ONTPD MedEd portal to have a standardized assessment form: • Redirection of care • Antenatal counseling	• Faculty and fellows felt this was an outstanding use of time at start of fellowship • Expanded to include PAs and NPs • Now incorporated as an annual event and writing new cases
First night-onCall: Experiential patient safety orientation for all incoming interns	• Transition from medical school difficulty for learner and institution and patients • Need to have common knowledge about institution-specific approach to safety issues and consolidation of medical knowledge prior to start of internship	**Learner**: All incoming interns from all programs (200) **Frequency** annual, during orientation **Feedback**: Faculty observer feedback at each station and group debriefs **Location**: Simulation center, in-person **Preparation**: Just-in-time WiseOnCall learning modules	**Number of cases**: 2 group OSCE (GOSCE) cases and 2 group simulations **Core topics:** • Environmental patient safety issues • Escalation • Medical error • Personal protective equipment • Structured handoffs	• Creates shared mental model for patient safety:1400+ interns have participated over 7 years, entire house staff now with common curriculum • Timing is key—During orientation so interns are eager and engaged • Electronic resource guide aligned with activities can be saved to their smart phone or iPad • Helped institution with ACGME and learning environment goals
Night-onCall: Near-graduation assessment of all 13 EPA prior to residency	• Assesses nearly graduated medical students' readiness for residency • Immersive 3-hour simulation includes a series of OSCE cases and activities that students will be expected to do during their first night on call as a resident • Goal is to help schools have tools to identify students individual learning plan for the start of residency	**Learner**: 340 near-graduating medical students in 2022 participated in 3-hour immersive simulation **Frequency:** Annual, near end of medical school **Feedback**: Learners receive detailed summary report and debrief with faculty **Location**: Simulation center, in-person or virtual	**Number of cases**: 3 cases and 3 activities **Core topics:** All 13 EPAs Medical learners are assessed on their communication, history and physical examination skills, note writing, culture of safety, clinical reasoning, transfer of case content to others, and evidence-based skills	• Funded by the Macy foundation to create a national NOC consortium • Now 2064 near-graduates' data from 9 schools will benchmark performance on a national level • Provides both learner assessment and medical school programmatic assessment • Potential for educational learning plan handoff to residency programs

Conclusion

The scope, complexity, and impact of the OSCEs described in this chapter illustrate that performance-based assessment is truly a team sport with the potential to affect our entire learning health system. Learners and teams can efficiently and effectively practice new skills or demonstrate competency in a simulation center (with an SP), remotely or in their own clinical environment (with ASPs or USPs). All these programs provide feedback at the individual, programmatic, and institutional levels either with verbal feedback, individual or aggregated reports, or even longitudinal dashboards. Faculty who participate as observers or case designers experience a new appreciation of their learner's competence and entrustment whether it is addressing knowledge, skills, or attitudes of systems-based practice, counseling on a new public health crisis, or demonstrating core communication skills when breaking bad news. This rigorous, nimble, tried, and true instructional design ensures that learning happens and entrustment is assessed. This is how we should continue training the next generation of clinicians and health systems.

References

Section 1: OSCEs to Teach New Curricular Topics

Virtual Heath at the Veterans Association Ambulatory Clinic

1. Phillips Z, Wong L, Crotty K, et al. Implementing an experiential telehealth training and needs assessment for residents and faculty at a veterans affairs primary care clinic. J Grad Med Educ. 2023;15(4):456–62. https://doi.org/10.4300/JGME-D-22-00868.1.

Addressing Discriminatory Patient Comments

2. Poitevien P, Osman C. Tackling implicit and explicit bias through objective structured teaching exercises for faculty. J Grad Med Educ. 2018;10(3):353–4. https://doi.org/10.4300/JGME-D-17-00906.1.
3. Torres C, Fagan I, Williams R, Porter B, Beltran C, Greene R, Phillips D, Agnant J, Pradhan D, Zabar S. From Student to intern, witness to upstander: Preparing incoming interns to respond to discriminatory patient behavior through simulation. Abstracts from the 2024 Annual Meeting of the Society of General Internal Medicine: J Gen Intern Med. 2024;39(2S) https://doi.org/10.1007/s11606-024-08797-5
4. Torres C, Morales D, Whitley A, Heller R, Horlick M, Porter B, Greene R, Zabar S. Co-producing an inclusive learning environment: residents and faculty design OSCEs to address bias by patients. Abstracts from the 2024 Annual Meeting of the Society of General Internal Medicine; J Gen Intern Med. 2024;39(2S) https://doi.org/10.1007/s11606-024-08797-5
5. Beltran CP, Wilhite JA, Gonzalez CM, Porter B, Torres C, Horlick M, Hauck K, Gillespie C, Zabar S, Greene RE. Requested a different doctor: developing and evaluating an OSCE assessing core skills in supporting trainees facing patient discrimination. J Gen Intern Med. 2025;40(1):207-212 https://doi.org/10.1007/s11606-024-09021-0.

Social Determinants of Health Screening in Ambulatory Care

6. Hanley K, Zabar S, Altshuler L, Lee H, Ross J, Rivera N, Marvilli C, Gillespie C. Opioid vs nonopioid prescribers: variations in care for a standardized acute back pain case. Subst Abus. 2017;38(3):324–9. https://doi.org/10.1080/08897077.2017.1319894.
7. Zabar S, Hanley K, Watsula-Morley A, Goldstein J, Altshuler L, Dumorne H, Wallach A, Porter B, Kalet A, Gillespie C. Using unannounced standardized patients to explore variation in care for patients with depression. J Grad Med Educ. 2018;10(3):285–91. https://doi.org/10.4300/JGME-D-17-00736.1.
8. Wilhite JA, Fisher H, Altshuler L, Cannell E, Hardowar K, Hanley K, Gillespie C, Zabar S. Gasping for air: measuring patient education and activation skillsets in two clinical assessment contexts. BMJ Simul Technol Enhanc Learn. 2020;7(5):428–30. https://doi.org/10.1136/bmjstel-2020-000759.
9. Gillespie C, Wilhite JA, Hanley K, Hardowar K, Altshuler L, Fisher H, Porter B, Wallach A, Zabar S. Addressing social determinants of health in primary care: a quasi-experimental study using unannounced standardised patients to evaluate the impact of audit/feedback on physicians' rates of identifying and responding to social needs. BMJ Qual Saf. 2022;27:bmjqs-2021-013904. https://doi.org/10.1136/bmjqs-2021-013904.
10. Altshuler L, Fisher H, Wilhite J, Phillips Z, Holmes I, Greene RE, Wallach AB, Smith R, Hanley K, Schwartz MD, Zabar S. Comparison of primary care patients' and unannounced standardized patients' perceptions of care. J Patient Exp. 2023;10:23743735231158940. https://doi.org/10.1177/23743735231158940.

Palliative Care Fellowship

11. Rau ME, Cohen SE. SIMplifying palliative extubations: a two-part osce in a simulation center to assess entrustable professional activities (EPAs) in hospice and palliative medicine (TH315). JPSM. 2020;59(2):413–4.

Emergency Medicine Faculty: Starting Buprenorphine in the Emergency Department

12. Lugassy D, McCormack RP, Shin SM, Zabar S, Ngyuen A, Moran Z, DiSalvo P. Demystifying and destigmatizing emergency department-initiated buprenorphine: A standardized experiential faculty development. Abstracts from the 2020 Annual Meeting of the Society of Academic Emergency Medicine. 2020; EMBASE:632417983 ISSN: 1553-2712 CID: 4547952

Telemedicine Care at Student Health Center

13. Phillips Z, Mitsumoto J, Fisher H, Wilhite J, Hardowar K, Robertson V, Paige J, Shahroudi J, Albert S, Li J, Hanley K, Gillespie C, Altshuler L, Zabar S. Using unannounced standardized patients to assess clinician telehealth and communication skills at an urban student health center. J Adolesc Health. 2024;74(5):1033-1038. https://doi.org/10.1016/j.jadohealth.2024.01.014.

Assessing COVID-19 Preparedness

14. Fisher H, Re C, Wilhite J, Hanley K, Altshuler L, Schmidtburger J, Gagliardi M, Zabar S. A novel method of assessing clinical preparedness for COVID-19 and other disasters. Int J Qual Health Care. 2021;33(1):mzaa116. https://doi.org/10.1093/intqhc/mzaa116.

Section 2: OSCEs to Teach/Assess Skills "Everyone" Needs

Care for Transgender and Gender Diverse (TGD) People

15. Greene RE, Garment AR, Avery A, Fullerton C. Transgender history taking through simulation activity. Med Educ. 2014;48(5):531–2. https://doi.org/10.1111/medu.12439.
16. Greene RE, Hanley K, Cook TE, Gillespie C, Zabar S. Meeting the primary care needs of transgender patients through simulation. J Grad Med Educ. 2017;9(3):380–1. https://doi.org/10.4300/JGME-D-16-00770.1.
17. Greene RE, Blasdel G, Cook TE, Gillespie C. How do OSCE cases activate learners about transgender health? Acad Med. 2020;95:S156–62. (12S Addressing Harmful Bias and Eliminating Discrimination in Health Professions Learning Environments) https://doi.org/10.1097/ACM.0000000000003704.
18. Dubin S, Kutscher E, Nolan I, Levitt N, Cook TE, Greene RE. Assessment of medical education on transgender health: a scoping literature review. Eval Health Prof. 2024;47(3):296–303. https://doi.org/10.1177/01632787231214531.
19. Beltran CP, Wilhite JA, Hayes RW, LoSchiavo C, Crotty K, Adams J, Hauck K, Crowe R, Kudlowitz D, Katz K, Gillespie C, Zabar S, Greene RE. Practice makes perfect: objective structured clinical examinations across the UME-to-GME continuum improve care of transgender simulated patients. J Grad Med Educ. 2024;16(2):182–94. https://doi.org/10.4300/JGME-D-23-00573.1. Epub 2024 Apr 15. PMID: 38993302; PMCID: PMC11234318

OBGYN Residency

20. Winkel AF, Lerner V, Zabar SR, Szyld D. A simple framework for assessing technical skills in a resident observed structured clinical examination (OSCE): vaginal laceration repair. J Surg Educ. 2013;70(1):10–4. https://doi.org/10.1016/j.jsurg.2012.08.005.
21. Winkel AF, Gillespie C, Hiruma MT, Goepfert AR, Zabar S, Szyld D. Test of integrated professional skills: objective structured clinical examination/simulation hybrid assessment of obstetrics-gynecology residents' skill integration. J Grad Med Educ. 2014;6(1):117–22. https://doi.org/10.4300/JGME-D-13-00055.1.
22. Winkel AF, Niles P, Lerner V, Zabar S, Szyld D, Squires A. Notes from the field: residents' perceptions of simulation-based skills assessment in obstetrics and gynecology. Eval Health Prof. 2016;39(1):121–5. https://doi.org/10.1177/0163278714563601.
23. Cardamone S, Vieira D, Winkel A. 66. Assessment of resident use of evidence-based practice in pediatric and adolescent gynecology utilizing objective structured clinical examination (OSCE). J Pediatr Adolesc Gynecol. 2020;33(2):206–7.

Reproductive Endocrinology Fellowship

24. Shaw J, Blatt B, Drda V, Zabar SR, Licciardi F. Simulation in reproductive endocrinology and infertility (REI): closing the counseling gap. Abstracts from the 78th Scientific Congress of the American Society for Reproductive Medicine; Fertil Steril. 2022;118(4S):e71.

Research Assistants Recruiting Older Adults to Clinical Trials

25. Fisher H, Zabar S, Chodosh J, Langford A, Trinh-Shevrin C, Sherman S, Altshuler A. A novel simulation-based approach to training for recruitment of older adults to clinical trials. BMC Med Res Methodol. 2022;22(1):180. https://doi.org/10.1186/s12874-022-01643-4.

Transitions of Care Management Skills

26. Heller R, Hillier E, Sartori D, Zabar S, Hayes R. Missing what's right before their eyes (on the screen): residents' postdischarge skills in a telemedicine visit with a standardized patient. Abstracts from the 2024 Annual Meeting of the Society of General Internal Medicine; J Gen Intern Med. 2024;39(2S). https://doi.org/10.1007/s11606-024-08797-5
27. Hillier E, Heller R, Sartori D, Zabar S, Hayes R. Taking note: assessing residents' medical chart documentation during a postdischarge telemedicine visit with a standardized patient. abstracts from the 2024 Annual Meeting of the Society of General Internal Medicine; J Gen Intern Med. 2024;39(2S). https://doi.org/10.1007/s11606-024-08797-5

Section 3: OSCEs as Programs of Assessment—Monitoring your Clinicians' Skills Development

Neurology Residency

28. Rostanski SK, Kurzweil AM, Zabar S, Balcer LJ, Ishida K, Galetta SL, Lewis A. Education research: simulation training for neurology residents on acquiring tPA consent: An educational initiative. Neurology. 2018;91(24):e2276–9. https://doi.org/10.1212/WNL.0000000000006651.

29. Valentine D, Kurzweil A, Zabar S, Lewis A. Objective Structured Clinical Exams (OSCE) are a feasible method of teaching how to discuss a nonepileptic seizure diagnosis. Epilepsy Behav. 2019;100(Pt A):106526. https://doi.org/10.1016/j.yebeh.2019.106526.
30. Kurzweil AM, Lewis A, Pleninger P, Rostanski SK, Nelson A, Zhang C, Zabar S, Ishida K, Balcer LJ, Galetta SL. Education research: teaching and assessing communication and professionalism in neurology residency with simulation. Neurology. 2020;94(5):229–32. https://doi.org/10.1212/WNL.0000000000008895.
31. Stainman RS, Lewis A, Nelson A, Zabar S, Kurzweil AM. Training in neurology: identifying and addressing struggling colleagues in the era of physician burnout. Neurology. 2020;95(17):796–9. https://doi.org/10.1212/WNL.0000000000010601.
32. Carroll E, Nelson A, Kurzweil A, Zabar S, Lewis A. Using Objective Structured Clinical Exams (OSCE) to teach neurology residents to disclose prognosis after hypoxic ischemic brain injury. J Stroke Cerebrovasc Dis. 2021;30(7):105846. https://doi.org/10.1016/j.jstrokecerebrovasdis.2021.105846.
33. LaRocque JJ, Grossman SN, Kurzweil AM, et al. Training in neurology: objective structured clinical examination case to teach and model feedback skills in neurology residency. Neurology. 2022;98(16):684–9. https://doi.org/10.1212/WNL.0000000000200223.
34. Zhang C, Kurzweil A, Pleninger P, Nelson A, Gurin L, Zabar S, Galetta SL, Balcer LJ, Lewis A. Neurology faculty comfort and experience with communication skills. J Clin Neurosci. 2023;109:21–5. https://doi.org/10.1016/j.jocn.2022.11.013.
35. Zakin E, Abou-Fayssal N, Lord AS, et al. Curriculum innovation: a standardized experiential simulation curriculum equips residents to face the challenges of chief year. Neurol Educ. 2024;3(3) https://doi.org/10.1212/ne9.0000000000200138.

Gastroenterology Fellowship

36. Chander B, Kule RR, Baiocco PJ, Chokhavatia S, Kotler D, Poles MA, Zabar S, Gillespie C, Ark T, Weinshel EH. Teaching the competencies: educational performance improvement using OSCEs for gastroenterology fellows. Clin Gastroenterol and Hepatol. 2009;7(5):509–14.
37. Alevi D, Baiocco P, Chokhavatia S, Kotler D, Poles M, Zabar S, Gillespie C, Ark T, Weinshel E. Teaching the competencies: using observed structured clinical examinations for faculty development. Am J Gastroenterol. 2010;105(5):973–7. https://doi.org/10.1038/ajg.2010.27.
38. Williams R, Miler R, Shah B, Chokhavatia S, Poles M, Zabar S, Gillespie C, Weinshel E. Observing handoffs and telephone management in GI fellowship training. Am J Gastroenterol. 2011;106(8):1410–4. https://doi.org/10.1038/ajg.2011.107.
39. Shah B, Roy Miler R, Poles M, Zabar S, Gillespie C, Weinshel E, Chokhavatia S. Informed consent in the older adult: OSCEs for assessing fellows' ACGME and geriatric gastroenterology competencies. Am J Gastroenterol. 2011;106(9):1575–9. https://doi.org/10.1038/ajg.2011.124.
40. Kingsbery J, Wolff M, Zabar S, Gillespie C, Weinshel E, Soloman A, Malter L. Optimizing transition of pediatric to adult care in inflammatory bowel disease (IBD) through the use of an observed structured clinical examination (OSCE). Am J Gastroenterol. 2014;109:S498–8.
41. Balzora S, Abiri B, Wang XJ, McKeever J, Poles M, Zabar S, Gillespie C, Weinshel E. Assessing cultural competency skills in gastroenterology fellowship training. World J Gastroenterol. 2015;21(6):1887–92. https://doi.org/10.3748/wjg.v21.i6.1887.
42. Wolff MJ, Balzora S, Poles M, Zabar S, Mintah A, Wong L, Weinshel E, Malter LB. Objective structured clinical examination as a novel tool in inflammatory bowel disease fellowship education. Inflamm Bowel Dis. 2015;21(4):759–65. https://doi.org/10.1097/MIB.0000000000000317.
43. Solomon AB, Reed R, Benkov K, et al. Using the objective structured clinical examination to assess ACGME competencies in pediatric gastroenterology fellows. J Pediatr Gastroenterol Nutr. 2017;64(4):e92–5. https://doi.org/10.1097/MPG.0000000000001450.
44. Chaudhary N, Lucero C, Villanueva G, Poles M, Gillespie C, Zabar S, Weinshel E. Assessment of abilities of gastroenterology fellows to provide information to patients with liver disease. Clin Gastroenterol and Hepatol. 2017;15(7):1095–123. https://doi.org/10.1016/j.cgh.2016.12.037.
45. Srisarajivakul N, Lucero C, Wang XJ, Poles M, Gillespie C, Zabar S, Weinshel E, Malter L. Disruptive behavior in the workplace: challenges for gastroenterology fellows. World J Gastroenterol. 2017;23(18):3315–21. https://doi.org/10.3748/wjg.v23.i18.3315.
46. Sartori DJ, Olsen S, Weinshel E, Zabar SR. Preparing trainees for telemedicine: a virtual OSCE pilot. Med Educ. 2019;53(5):517–8. https://doi.org/10.1111/medu.13851.
47. Papademetriou M, Perrault G, Pitman M, Gillespie C, Zabar S, Weinshel E, Williams R. Subtle skills: Using objective structured clinical examinations to assess gastroenterology fellow performance in system based practice milestones. World J Gastroenterol. 2020;26(11):1221–30. https://doi.org/10.3748/wjg.v26.i11.1221.

Primary Care Residency Program and Internal Medicine Residency Program

48. Zabar S, Hanley K, Stevens D, Kalet A, Schwartz M, Pearlman E, Kachur BJ, E, Lipkin M. Measuring the competence of residents as teachers. J Gen Intern Med. 2004;19(5 Pt 2):530–3. https://doi.org/10.1111/j.1525-1497.2004.30219.x.
49. Zabar S, Hanley K, Kachur E, Stevens D, Schwartz M, Pearlman E, Adams J, Felix K, Lipkin M, Kalet A. "Oh! She doesn't speak English!" Assessing resident competence in managing linguistic and cultural barriers. J Gen Intern Med. 2006;21(5):510–3. https://doi.org/10.1111/j.1525-1497.2006.00439.x.
50. Zabar S, Hanley K, Adams J, Ark TK. No place like home: teaching home visits. Med Educ. 2009;43(11):1102–3. https://doi.org/10.1111/j.1365-2923.2009.03487.x.
51. Zabar S, Adams J, Kurland S, Shaker-Brown A, Porter B, Horlick M, Hanley K, Altshuler L, Kalet A, Gillespie C. Charting a key competency domain: understanding resident physician interprofessional collaboration (IPC) skills. J Gen Intern Med. 2016;31(8):846–53. https://doi.org/10.1007/s11606-016-3690-6.
52. Zabar S, Hanley K, Horlick M, Cocks P, Altshuler L, Watsula-Morley A, Berman R, Hochberg M, Phillips D, Kalet A, Gillespie C. "I Cannot Take This Any More!": preparing interns to identify and help a struggling colleague. J Gen Intern Med. 2019;34(5):773–7. https://doi.org/10.1007/s11606-019-04886-y.
53. Lawrence K, Hanley K, Adams J, Sartori DJ, Greene R, Zabar S. Building telemedicine capacity for trainees during the novel coronavirus outbreak: a case study and lessons learned. J Gen Intern Med. 2020;35(9):2675–9. https://doi.org/10.1007/s11606-020-05979-9.
54. Sartori DJ, Hayes RW, Horlick M, Adams JG, Zabar SR. The TeleHealth OSCE: preparing trainees to use telemedicine as a tool for transitions of care. J Gen Intern Med. 2020;1;12(6)764–8. https://doi.org/10.4300/JGME-D-20-00039.1.
55. Wilhite JA, Zabar S, Gillespie C, Hauck K, Horlick M, Greene RE, Hanley K, Adams J. "I don't trust it": use of a routine OSCE

to identify core communication skills required for counseling a vaccine-hesitant patient. J Gen Intern Med. 2022;37(9):2330–4. https://doi.org/10.1007/s11606-022-07495-4.
56. Beltran CP, Wilhite JA, Hayes RW, LoSchiavo C, Crotty K, Adams J, Hauck K, Crowe R, Kudlowitz D, Katz K, Gillespie C, Zabar S, Green R. Practice makes perfect: objective structured clinical examinations across the UME-to-GME continuum improve care of transgender simulated patients. J Grad Med Educ. 2024;16(2):182–94. https://doi.org/10.4300/JGME-D-23-00573.1.
57. Beltran CP, Wilhite JA, Gonzalez CM, Porter B, Torres C, Horlick M, Hauck K, Gillespie C, Zabar S, Greene RE. Requested a different doctor: developing and evaluating an OSCE assessing core skills in supporting trainees facing patient discrimination. J Gen Intern Med. 2025;40(1):207–12. https://doi.org/10.1007/s11606-024-09021-0.

Pediatric Residency

58. Howell HB, Desai PV, Altshuler L, McGrath M, Ramsey R, Vrablik L, Levy FH, Zabar S. Teaching and assessing communication skills in pediatric residents: how do parents think we are doing? Acad Pediatr. 2022;22(2):179–83. https://doi.org/10.1016/j.acap.2021.06.011.
59. Desai PV, Howell HB, McGrath M, Ramsey R, Lebowitz J, Trogen B, Cha C, Pierce KA, Zabar S. Zoom objective structured clinical exams: virtually the same as the real thing? *Acad Pediatr*. 2023 23(2):483–488. https://doi.org/10.1016/j.acap.2022.11.004.

Surgery Residency

60. Hochberg M, Kalet A, Zabar S, Kachur E, Gillespie C, Berman R. Can professionalism be taught? Encouraging evidence. Am J Surg. 2010;199(1):86–93. https://doi.org/10.1016/j.amjsurg.2009.10.002.
61. Hochberg M, Berman R, Kalet A, Zabar S, Gillespie C, Pachter H. The professionalism curriculum as a cultural change agent in surgical residency education. Am J Surg. 2012;203(1):14–20. https://doi.org/10.1016/j.amjsurg.2011.05.007.
62. Hochberg M, Berman R, Kalet A, Zabar S, Gillespie C, Pachter L. The stress of residency: recognizing the signs of depression and suicide in you and your fellow students. Am J Surg. 2013;205(2):141–6. https://doi.org/10.1016/j.amjsurg.2012.08.003.
63. Hochberg MS, Berman RS, Kalet AL, Zabar S, Gillespie C, Pachter HL. Professionalism training for surgical residents: documenting the advantages of a professionalism curriculum. Ann Surg. 2016;264(3):501–7. https://doi.org/10.1097/SLA.0000000000001843.

Section 4: OSCEs to Prepare Learners for Transitions

Onboarding New Faculty

64. Wallach A, McCrickard M, Eliasz KL, Hochman K. An experiential faculty orientation to set communication standards. *Med Educ*. 2019 53(5):512–513. https://doi.org/10.1111/medu.13867.
65. Heller R, Wilhite J, Phillips Z, Wallach A, Hochman K, Zabar S. Mpox simulation: a nimble educational tool during an emergent public health outbreak. Abstracts from the 2024 Annual Meeting of the Society of General Internal Medicine; J Gen Intern Med. 2023;38(S2). https://doi.org/10.1007/s11606-023-08226-z
66. Heller R, Wilhite J, Hochman K, Wallach A, Zabar S. Clearing the Air: Counseling on the Air Quality Index (AQI) and preventing asthma exacerbations during periods of poor air quality. Abstracts from the 2024 Annual Meeting of the Society of General Internal Medicine; J Gen Intern Med. 2024;39(2). https://doi.org/10.1007/s11606-024-08797-5

Anesthesiology Residency Program

67. Chanan E, Rollins MD. Adding to Our Competitive Advantage: Making the Case for Teaching Communication and Professionalism. Anesthesia & Analgesia. 2018;127(2):580–5. https://doi.org/10.1213/ANE.0000000000002737.

Neonatal Intensive Care Fellows

68. Angert R. Monitoring of Central Catheters by Neonatal Learners-Effects of an Objective Driven Ultrasound Training Curriculum. Presented at the 97th Perinatal and Developmental Medicine Symposium in June 2021.

First Night-onCall: Experiential Patient Safety Orientation for All Incoming Interns

69. Eliasz KL, Kalet A, Buckvar-Keltz L, Phillips D, Riles TS, Manko J, Ng GM, Andrade GN, Zabar S. A culture of safety from day 1: an institutional patient safety initiative to support incoming interns. JGME. 2018;10(3):347–9. https://doi.org/10.4300/JGME-D-17-00828.1.
70. Beltran CP, Pradhan D, Phillips D, Agnant J, Tame K, Blatt B, Buckvar-Keltz L, Manko JA, Zabar S. Are incoming interns prepared for patient safety? New York University Grossman School of Medicine 5-year experience. Acad Med. 2023;98(11S):S193–4. https://doi.org/10.1097/ACM.0000000000005350.

Night-onCall: Near-Graduation Assessment of All 13 EPAs Prior to Internship

71. Kalet A, Zabar S, Szyld D, Yavner SD, Song H, Nick MW, Ng G, Pusic MV, Denicola C, Blum C, Eliasz KL. A simulated "Night-onCall" to assess and address the readiness-for-internship of transitioning medical students. Adv Simul (Lond). 2017;2:13. Published 2017 Aug 14. https://doi.org/10.1186/s41077-017-0046-1.
72. Eliasz KL, Nick MW, Zabar S, Buckvar-Keltz L, Ng GM, Riles TS, Kalet AL. Viewing Readiness-for-Residency through Binoculars: Mapping Competency-Based Assessments to the AAMC's 13 Core Entrustable Professional Activities (EPAs). Teach Learn Med. 2022;6:1–6. https://doi.org/10.1080/10401334.2022.2082432.
73. Eliasz KL, Ark TK, Nick MW, Ng GM, Zabar S, Kalet AL. Capturing entrustment: using an end-of-training simulated workplace to

assess the entrustment of near-graduating medical students from multiple perspectives. Med Sci Educ. 2018;28(4):739–47. https://doi.org/10.1007/s40670-018-0628-0.

74. Chang LY, Eliasz KL, Cacciatore DT, Winkel AF. The transition from medical student to resident: a qualitative study of new residents' perspectives. Acad Med. 2020;95(9):1421–7. https://doi.org/10.1097/ACM.0000000000003474.
75. Ark T, Wargo E, Crowe R, DeWitt D, Auerbach L, Schaye V, Dodson L, Frencis M, Kalet A, Zabar S. How ready are our near-graduates for internship? Night-onCall Immersive Simulation Data over time and across consortium schools. Abstracts from the 2023 Annual Meeting of the Society of General Internal Medicine; J Gen Intern Med. 2023a;38(2). https://doi.org/10.1007/s11606-023-08226-z
76. Ark T, Wargo E, Zabar S, Kalet A. How ready are our near-graduates to be on call? Some are and some aren't. Abstracts from the 2023 Annual Meeting of the Society of General Internal Medicine; J Gen Intern Med. 2023b;38(2). https://doi.org/10.1007/s11606-023-08226-z

9 Simulating Virtual Care: Integrating Telemedicine into Objective Structured Clinical Training

Daniel J. Sartori and Katharine Lawrence

The COVID-19 pandemic catalyzed a dramatic transition toward virtual care, launching a "new normal" of telemedicine-based care for clinical practices, including training residents and medical students. As a result, telemedicine skills have become an important part of the medical learner's foundational clinical skillset. Until very recently, telemedicine-specific skills required for successful video and phone encounters were not emphasized in undergraduate or graduate medical education (UME, GME). While didactic curricula and clinical telemedicine experiences currently exist in several medical schools [1, 2] and GME training programs [3, 4], telemedicine curricula and structured assessment tools are still in their infancy. In 2021, the Association of American Medical Colleges (AAMC) made an initial effort to address this critical need by developing telehealth training competencies to support curricular and professional development across the learning continuum and guide competency-based education (CBE) for telemedicine [5, 6].

As the world of telemedicine-based healthcare expands, it is crucial that learners have opportunities to develop their skills in a well-structured and supportive environment that provides opportunities for meaningful learning, assessment, and feedback. Tools are emerging to provide these opportunities, including simulating video-based care through telemedicine OSCEs (also known as "tele-OSCEs," telemedicine or telehealth OSCEs, or virtual OSCEs) and virtual workplace-based assessments [7, 8]. Increasingly, training programs are taking advantage of virtual health simulations to introduce learners to this modality and assess virtual clinical skills, as well as explore new domains in telemedicine care such as multidisciplinary virtual care coordination [9], virtual triage [10], virtual urgent care [11], and more. Medical educators should consider incorporating virtual health simulations into their curricula.

D. J. Sartori (✉)
Department of Medicine, New York University Grossman School of Medicine, New York, NY, USA
e-mail: Daniel.Sartori@nyulangone.org

K. Lawrence
Department of Population Health, New York University Grossman School of Medicine, New York, NY, USA

Building Virtual Health Simulations: Competencies, Cases, and Checklists

Virtual health simulations can teach and evaluate telemedicine skills at two levels: (1) telemedicine-specific skills (e.g., technical competence with device setup, technology troubleshooting, remote-specific physical examination) and (2) core clinical skills that are mediated through the virtual environment, such as interpersonal communication and professionalism. A well-structured simulation can enhance the core clinical skills of the learner as well as the relevant context-specific skills of the new telemedicine environment. These simulations can be conducted through commercial virtual conferencing tools (e.g., Zoom™, FaceTime™) or the electronic health record (EHR) system/portal your healthcare system uses.

Faculty should select cases for simulated virtual encounters that reflect the visit types best suited to this modality. In particular, virtual care is most useful for chronic symptom surveillance, real-time urgent evaluation, and follow-up for vulnerable patients, such as those recently discharged from the hospital. Our group has developed internal medicine-specific cases assessing virtual management of chronic diseases [8, 12], transitions of care from hospital to community [13], and just-in-time virtual urgent care [11, 14], with the intent of assessing clinical management, core communication, and telemedicine skills. We encourage others to design cases to simulate other high-yield virtual visit types to provide the most authentic training for learners.

Until recently, no formalized virtual health competencies existed to guide development of virtual health education for learners. However, the introduction of AAMC telehealth competencies across the training continuum, as well as the incorporation of "Digital Health" in the ACGME's Milestones 2.0 for Internal Medicine, provided significant

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_9

guidance for training and assessment of learners. Telehealth competency domains identified by the AAMC include (1) patient safety and appropriate use, (2) access and equity, (3) communication, (4) data collection and assessment, (5) technology, and (6) ethical practices and legal requirements. The AAMC establishes specific skills in each of these domains and provides targets for each of the following levels of learners: those recently entering residency, those starting clinical practice, and those with experienced clinical practice (3–5 years post-residency). Guides, including formal competencies, are available for both UME- and GME-level medical trainees, as well as practicing physicians, nurses, and allied health professionals engaging in telehealth in clinical settings [15–17].

Virtual health simulations can be reliably assessed by a standardized patient (SP) using a behaviorally anchored checklist developed to assess the competencies. Table 9.1 shows how our telemedicine skills map to the AAMC competencies. Specifically, this demonstrates how a "well-done" telemedicine skill maps to the AAMC competency for "entering practice," given that our checklist is primarily used for resident-level learners. In high-stakes assessments, this

Table 9.1 Telemedicine assessment items and AAMC competency domains

	Assessment of item			
Telemedicine skill	Not done	Partially done	Well done	AAMC competency domain
Partnered with patient to perform virtual physical exam	Made no attempt to perform physical exam	Asked patient to perform basic exam maneuvers or utilize peripheral monitoring devices (thermometer, home BP cuff, Fitbit/Apple watch, etc.)	Asked patient to perform maneuvers or access peripheral monitoring device followed by verbal confirmation of findings with patient or collateral	Domain 4: Data collection and assessment via telehealth 4.2.b Conducts appropriate physical examination or collects relevant data on clinical status during a real or simulated telehealth encounter, including guiding the patient and/or tele-presenter
Confirmed patient identifiers	Did not attempt to identify patient or patient's location	Asked patient to confirm either (a) name and/or date of birth, (b) call back number, OR (c) location	Asked patient to confirm (a) name and/or date of birth, (b) call back number, AND (c) location	Domain 6: Ethical practices and legal requirements for telehealth 6.1.b Complies with legal and privacy regulations for telehealth at the local, state, and federal levels
Exhibited comfort and confidence using video interface	Was shy/uncomfortable in front of the camera, let technological glitches (if any) distract from the encounter	Mostly comfortable on camera, occasional stumble but interview was not derailed by occasional video delay or glitch	Confident on camera, acknowledged and moved forward from technical glitches, and did not let video interface detract from natural conversation	Domain 3: Communication via telehealth 3.2b Establishes therapeutic relationships and environments during video visits, such as attending to disruptions related to privacy, lighting, sound, and attire
Used nonverbal communication to enrich communication on camera	Avoided eye contact via the webcam, slouched, or was angled away or too far from camera	Made occasional eye contact with webcam, sometimes slouched or out of view	Maintained eye contact with webcam throughout encounter, sat squarely in front of camera, and at appropriate distance	Domain 3: Communication via telehealth 3.1.b. Develops an effective rapport with patients via video visits, attending to eye contact, tone, body language, and nonverbal cues
Utilized live video to augment information gathering	Made no attempt to visually reconcile medications, witness reproducible symptoms, talk with onsite collateral (family/HHA/VNS), visual tour of home	Did one of the following: Visually reconcile medications, witness reproducible symptoms, talk with onsite collateral (family/HHA/VNS), visual tour of home	Did two or more of the following: Visually reconcile medications, witness reproducible symptoms, talk with onsite collateral (family/HHA/VNS), visual tour of home	Domain 3: Communication via telehealth 3.3.b Determines situations in which patients' social supports and health care providers should be incorporated into telehealth interactions, with the patients' consent, to provide optimal care

(continued)

Table 9.1 (continued)

	Assessment of item			
Telemedicine skill	Not done	Partially done	Well done	AAMC competency domain
Actively optimized technical aspects of the virtual encounter	Did not assess sound (e.g., volume, clarity, background noise), video (e.g., pixilation or delay), or "backup plan" if technology failed (e.g., phone call)	Assessed two of the following: Sound (e.g., volume, clarity, background noise), video (e.g., pixilation or delay), or "backup plan" if technology failed (e.g., phone call)	Assessed three of the following: Sound (e.g., volume, clarity, background noise), video (e.g., pixilation or delay), or "backup plan" if technology failed (e.g., phone call)	Domain 3: Communication via telehealth 3.2.b. Establishes therapeutic relationships and environments during video visits, such as attending to disruptions related to privacy, lighting, sound, and attire Domain 5: Technology for telehealth 5.3.b Demonstrates how to troubleshoot basic technology failures and optimize settings with the technology being used
Maintained appropriate computer etiquette during encounter	Frequent typing without explanation, appears preoccupied with computer, or distracted	Occasional types with minimal explanation of actions	Paused video or provided clear explanation while documenting, searching another website, or having another screen open for the purpose of patient care	Domain 3: Communication via telehealth 3.1b Develops an effective rapport with patients via video visits, attending to eye contact, tone, body language, and nonverbal cues

checklist can be completed by the SP post-encounter; for lower-stakes formative exercises, the checklist can be used by an observing faculty member or group of learners to provide structured feedback. Note that while our checklists separate telemedicine skills as a distinct domain for clarity (these include virtual patient identification, use of video to augment history gathering, optimizing audio/video interface, performing a patient-centered virtual physical exam, and maintenance of "webside" manner), we acknowledge that these skills overlap with general core communication domains expected of all trainees in all patient encounters. The checklist items are agnostic to specialty or visit type and enable assessment in and across a variety of contexts. Sample telemedicine assessment items that can be added to the SP checklist for a virtual telehealth OSCE are presented in Appendix R.

Optimizing Your Environment for Telemedicine OSCEs: For Educators and Learners

To optimize the educational value of a virtual health simulation, it is critical to create realistic scenarios and settings. When using a commercial conference tool, the learner and SP can be remote (at home), onsite (in the simulation center), or a combination of both. Incorporating the live EHR into simulated visits, by using its integrated video software and visit documentation, can add to the authenticity of the encounter. In these cases, we recommend working with your information technology (IT) or EHR vendor partners to set up a simulation experience that is close to real-life environments.

Fully remote telemedicine OSCEs—where the learner, SP, and potentially a faculty member are all accessing the case from different remote locations—are viable alternatives for institutions where robust OSCE infrastructure (e.g., simulation "SIM" centers) is not available, local expertise is limited (e.g., specialty care evaluations such as tele-neurology), or learners are dispersed. Important considerations for fully remote telemedicine OSCEs are included in Table 9.2. These include remote security clearances for assessing institutional technology off-site (if not using conference-based applications) and accounting for the individual participant's hard- and software, including Internet connectivity. In addition, it is important that educators work to create inclusive environments for their OSCE participants, particularly around technology accessibility. A brief survey of accessibility needs prior to the initiation of OSCE planning may suffice, or more robust efforts can be made to ensure a minimum level of universal accessibility for all telemedicine-based training encounters. More information on how to achieve this is available via the HHS website Guidance on Nondiscrimination in Telehealth [18]. The SP's technology skills and accessibility must also be assessed and addressed.

Key considerations regarding the simulation environment, the hard- and software, and the SP training are outlined below; we recommend that planning include a thorough

Table 9.2 Practical considerations for educators interested in developing tele-OSCEs with an existing telemedicine platform

Encounters where learner is off-site or in simulation center	• Can learners access EHR or web-based conference platform remotely? Do all learners have internet sufficient to support video? • Will encounter be recorded? If so, what additional equipment is needed to record (audio and/or visual) learners in remote locations? • How do we prepare for remote technical issues and provide remote technical support as needed (e.g., "zoom lag")? • What opportunities exist to incorporate remote experts from other locations or institutions into these experiences? • Are there scenarios where our OSCEs will be *fully remote* (e.g., learners, SPs, and evaluators are all at separate locations conducting the OSCE on a shared video conferencing site)? – If yes, how do we connect learners, SPs, and evaluators? What security authorizations are needed to remote access the technology platforms?.
Encounters where learner is in their clinical practice/ at clinical site	• Is the virtual simulation a telemedicine OSCE with a web-based conference platform or the institution's EHR? • If simulating with the EHR, can a mock patient chart be created within the EHR? What support is needed to accomplish this? • What, if any, special features do our platform offer (e.g., multi-party conferencing, live translation services)? How might we assess use of these features? • What permissions are needed for learners to access computers, EHRs, and/or the telemedicine platform? Who will obtain these? Who will ensure the OSCE stations have the required equipment? • What permissions are needed for SPs to access the EHR? • What software do we need to make the telemedicine encounter work? Examples: EHR, telemedicine vendor platform, middleware/third party apps, and security access. • (How) does the telemedicine platform integrate with these other programs? When are major software updates scheduled? How will these affect our OSCE? • Do all computers on site have microphones, speakers, and cameras? • What kind of information technology (IT) support is available? Who will be responsible for troubleshooting technical issues during the OSCE?.
SP training and support considerations	• Where will the SP be located during the OSCE? • What should the SP's telemedicine setup look like (e.g., phone-based, computer-based, etc.)? • How much technical training/familiarity should the SP have prior to the case? • What does the SP'S environment look like on screen? Do you want to simulate a hospital bed or a specific room in their house? Do you want a fake background? • Does SPs medical data need to be provided during the case? If so, how? Can the mock SP in the chart be prepopulated with demographics, medical history, allergies, meds, outpatient pharm, recent hospital notes, discharge summary/meds? • If there is no medical chart being used, is a "door note" (the case information usually posted on the door outside of the simulation center room to give context to the learner) needed? Who posts the door note if using a web-based conference platform? • How will the SP assess the learner? Learning management system (B-line/learning space/EMS) or other data collection systems (REDCap)?.
Additional considerations	• How do we support students and/or SPs with accessibility accommodations? • What other tools and/or props do we need for the case (e.g., pill bottles, remote monitoring devices)? • How will we make-up sessions if there are significant technical issues during the OSCE (e.g., system-wide outages)?

review and a pilot of all technology and cases prior to undertaking a virtual simulation.

Table 9.3 offers practical guidance for both learners and clinicians new to telemedicine practice. We developed it based on our experience and that of others [12, 19, 20] and use it for preparation or as a post-experience resource. These encounters offer learners the opportunity to practice technical aspects of the video (home setup, EHR logistics) and become comfortable with key components of the telemedicine encounter such as asking for two patient identifiers (as well as patient location, if applicable) and confirming a patient can see and hear appropriately for the visit.

In some cases, an OSCE dedicated purely to the optimization of the telemedicine clinical environment may be appropriate for new learners. For these cases, the training goal is developing comfort and facility with the technologies at hand, such as adapting audio and visual setups to improve ergonomics for the clinician or educating patients in optimizing their own environments (e.g., asking patients to move from a public area to a quieter space or to adjust their cameras so that their entire face is visible on screen). Such cases offer unique opportunities to assess learners on important interpersonal communication and patient education skills; other cases can help learners identify and address unique barriers to effective healthcare technology use for their patients such as digital health literacy, confidence, or trust [21].

Of note, during the OSCE, educators should provide clear guidance to both learners and SPs on what is in or out of

Table 9.3 Optimizing telemedicine environments for clinical encounters: a resource for preparation or post-experience debrief

Recommendations for learners: *Physical environment*	• Create a private, quiet, and professional space: – Ensure you are in a private space, without disruptions from roommates, family, etc. – Remove unnecessary clutter and personal items – Find a solid colored (preferably light) background to sit in front of • Wear appropriate professional attire: – Avoid bright colors, stripes, or bold patterns – Remove shiny jewelry, lapel pins, etc., as they may cause reflective artifacts onscreen – Be aware of any commercial or industry labels, logos, or identifying signage on your clothing or background
Recommendations for learners: *Audio/visual and computer setups*	• Ensure your device (computer, tablet, phone) is fully charged; close unnecessary programs or applications that may be running in the background • Consider logging in early to do a status check of your camera, microphone, and other devices • Minimize ambient noise—Close windows and/or doors, and consider using headphones (be sure to set computer settings to headphone audio and microphone) • Find a well-lit area—Front or overhead lighting is best; backlighting can obscure your image • Position yourself in a centered, "middle close-up" location on the screen; an ideal image includes a full view of your face and the upper half of your torso – If possible, adjust the location of the patient's image to be closest to your camera lens—this will give the impression that you are "looking" at the patient's face when you look at the screen and/or camera
Recommendations for learners: *Conducting the telemedicine visit*	• Aim to see patients on time; attempt to notify patients if delays in care are expected (note: Many patients will be placed in a virtual "waiting room" until their visit is started) – Speak clearly and in a normal voice; if you normally speak very quickly, consider slowing your speech to accommodate for sound quality issues, connectivity delays, or patient needs • Identify yourself and the patient! – Two patient identifiers (name, DOB) • Confirm the patient's location (if relevant for billing and/or licensing purposes), and obtain a call-back number • Ensure audio/visual connectivity; ask patients "can you hear and see me ok?" • Ensure the patient is in a private space and is able to share information freely • Introduce the patient to the virtual encounter, and let them know what to expect – *Benefits*: More convenient, flexible, allows access over long distances – *Limitations*: Not everything can be done in a virtual visit; if something is urgent or there is a serious health concern, a follow-up in-person visit (or ED referral) may be required • Determine the appropriateness of a virtual visit – Active chest pain? Complex or sensitive physical exam needs? Inability to acquire complete or accurate information to make a medical decision? Consider converting the patient to an in-person follow-up

scope of the evaluation of technical competencies during the simulation. For example, if learners are being evaluated on their ability to create an optimized telemedicine environment for their patients, you may or may not wish to advise the learner before the encounter that it is appropriate to adjust the A/V equipment so that both patients and providers can see and hear one another sufficiently. Warning learners that they are expected to actively interact with the technology and optimize technical issues as part of the OSCE guides learners to demonstrate their technical skills.

Special Considerations for Virtual Health Simulation

While scheduled virtual OSCEs provide a model for skills-based telemedicine training, they do so in a laboratory setting, removing learners from their larger clinical environment, imposing artificial constraints on visit length, and in many cases, substituting an artificial audio/visual interface for the one used in practice. Workplace-based virtual health simulation, however, has the potential to create a more authentic training experience by integrating assessment into learner's real-world clinical schedule and practice. Such simulated encounters can be "announced" in which learners are aware that the visit is simulated or "unannounced" in which virtual standardized patient encounters occur covertly. Both announced and unannounced workplace-based virtual encounters require considerable planning and partnership with local healthcare information technology to execute, given the challenges of pausing real-world clinical practice to implement a simulated patient care visit. Unannounced virtual visits add an additional layer of complexity given the need to create mock electronic patient records that closely mimic those from real patients and allow realistic documentation and orders to be placed, all while ensuring that normal clinical workflows are not disrupted. Despite these complexities, workplace-based virtual encounters can be powerful and authentic training tools with the potential for longitudinal assessment and more durable practice change. Table 9.4 compares and contrasts these models.

Table 9.4 Telemedicine-specific considerations for creating a work-place-based telemedicine simulation

Framework for selecting telemedicine assessment instruction design	Performance based assessment in simulation center or on web-based conference platform	Announced standardized patient in clinical setting	Unannounced standardized patient/ secret shopper in clinical setting
Encounter structure	Learner participates in simulated telehealth encounter outside of clinic schedule, through zoom or other teleconferencing system	Learner and clinic staff are informed that *a specific telemedicine visit* in their EHR schedule is a standardized patient	Learner is informed that a telemedicine visit in their EHR schedule will be a standardized patient, but *specific patient is not disclosed*
Location	Simulation center or zoom	Learner's clinical setting	Learner's clinical setting
Technology needs	Zoom or another video conferencing platform	Ability to integrate visit into clinical schedule; okay if patient does not appear realistic in EHR	Ability to generate mock patient and integrate visit into EHR and clinical schedule without detection
Feedback	• Oral feedback from SP and faculty observer immediately following visit • Written feedback	• Oral feedback from SP immediately following visit • Written feedback	• Written feedback from SP post hoc
Strengths of this structure/unique learning opportunities	(1) Simple to implement (2) Learners receive immediate feedback from faculty and SP	(1) Visit integrated into learner's workflow (2) Learner practices EHR-specific skills of their clinical environment (3) Learners receive immediate feedback from SP	(1) Visit integrated into learner's workflow (2) Learner practices EHR-specific skills of environment (3) Insight into clinical integration of telemedicine

The Future of Virtual Health Simulations

The future of clinical care will no doubt include virtual health, which presents challenges and opportunities for medical educators. Virtual healthcare has already changed care delivery and created new models of care continuity and patient-centered care while shifting expectations around accessibility and experience for both patients and providers. While the extent to which telemedicine is incorporated into current curricula is variable, enthusiasm for the practice is growing, and medical organizations and accrediting bodies are moving toward standardized training requirements, signaling their recognition of the role that telemedicine will play in learners' careers. Additional considerations for telemedicine OSCE skills include virtual oversight of learners ("virtual precepting"), as well as attention to documentation and billing. Advanced areas of exploration may include the incorporation of other digital health technologies such as remote patient monitoring (RPM), which can allow learners to experience scenarios and develop skills in both synchronous and asynchronous data collection, review, and clinical decision-making.

In the realm of virtual professional development, there is a need for more robust understanding of and training in concepts of digital professionalism, digital empathy, and digital trust among medical learners. These concepts—derived from the fields of social psychology, media studies, and computer science—encapsulate the adaptation of professional behavior, codes of conduct, and social norms into virtual spaces [22, 23] and explore the role of technology as a mediator and moderator of critical elements of the patient–physician relationship. Finally, with growing concern around the role of health technologies in creating or exacerbating health inequity, it is critical that the future of OSCEs in telemedicine education directly addresses equitable access to digital health resources [21, 24] and provides training to learners to actively identify and address patients' digital determinants of health. These represent new and exciting, though as yet understudied, avenues for the development of OSCEs geared toward ensuring the next generation of medical learners have the capacity to leverage emerging digital and virtual health technology in the care of individual patients and the public's health.

Best Practices for Telemedicine OSCEs

1. Select a simulation modality and setting that meets the needs of learners and your specific clinical environment.
2. Base telemedicine simulations and assessments on established competency frameworks, such as those proposed by the AAMC.
3. Seek to assess both general skills in the context of telemedicine, as well as telemedicine-specific skills.
4. Utilize behaviorally anchored assessment tools for evaluating learners.
5. Stay current with telemedicine trends in OSCE simulations; ensure that simulations reflect the most recent telemedicine technologies, practices, and regulations.

6. Simulate cases that are most appropriate for assessing telemedicine-based competencies.
7. Design OSCE simulation spaces to closely resemble real-world telemedicine environments; this includes, where possible, incorporating live or training versions of the current telemedicine and EHR platforms utilized by your health system.
8. If possible, partner with your health system's information technology department to support virtual simulation infrastructure.
9. Accommodate participants' technical constraints in simulations; understand the hard- and software setups of your participants (learners, SPs, and evaluators), and be prepared to accommodate those with access constraints (e.g., broadband limitations, security settings on computers).
10. Be prepared to adapt environments and technologies for learners who have accessibility and/or adaptive needs.
11. Incorporate digital health equity considerations in telemedicine simulations; design scenarios that account for structural, social, and digital determinants of health for your patients and that work to build digital health equity competencies in learners.
12. Evaluate digital professionalism and empathy skills in telemedicine simulation assessments.

References

1. Farrell SE, Junkin AR, Hayden EM. Assessing clinical skills via telehealth objective standardized clinical examination: feasibility, acceptability, comparability, and educational value. Telemed J E Health. 2022;28(2):248–57. https://doi.org/10.1089/tmj.2021.0094. Epub 2021 May 17. PMID: 33999715
2. Henschen BL, Jasti H, Kisielewski M, Pincavage AT, Levine D. Teaching telemedicine in the COVID-19 era: a national survey of internal medicine clerkship directors. J Gen Intern Med. 2021;36(11):3497–502. https://doi.org/10.1007/s11606-021-07061-4. Epub 2021 Sep 7. PMID: 34494207; PMCID: PMC8423074
3. Kirkland EB, DuBose-Morris R, Duckett A. Telehealth for the internal medicine resident: a 3-year longitudinal curriculum. J Telemed Telecare. 2021;27(9):599–605. https://doi.org/10.1177/1357633X19896683. Epub 2019 Dec 31. PMID: 31888396
4. Savage DJ, Gutierrez O, Montané BE, Singh AD, Yudelevich E, Mahar J, Brateanu A, Khatri L, Fleisher C, Jolly SE. Implementing a telemedicine curriculum for internal medicine residents during a pandemic: the Cleveland Clinic experience. Postgrad Med J. 2022;98(1161):487–91. https://doi.org/10.1136/postgradmedj-2020-139228. Epub 2021 Mar 10. PMID: 33692154
5. AAMC. Telehealth competencies across the learning continuum. AAMC new and emerging areas in medicine series. Washington, DC: AAMC; 2021.
6. Galpin K, Sikka N, King SL, Horvath KA, Shipman SA. AAMC telehealth advisory committee. Expert consensus: telehealth skills for health care professionals. Telemed J E Health. 2021;27(7):820–4. https://doi.org/10.1089/tmj.2020.0420. Epub 2020 Nov 25. PMID: 33236964
7. Martin R, Mandrusiak A, Russell T, Forbes R. A toolbox for teaching telehealth using simulation. Clin Teach. 2022;19(4):270–5. https://doi.org/10.1111/tct.13510. Epub 2022 Jun 21. PMID: 35726556; PMCID: PMC9543712
8. Phillips Z, Wong L, Crotty K, et al. Implementing an experiential telehealth training and needs assessment for residents and faculty at a veterans affairs primary care clinic. J Grad Med Educ. 2023;15(4):456–62. https://doi.org/10.4300/JGME-D-22-00868.1.
9. Wong R, Ng P, Spinnato T, Taub E, Kaushal A, Lerman M, Fernan A, Dainer E, Noel K. Expanding telehealth competencies in primary care: a longitudinal interdisciplinary simulation to train internal medicine residents in complex patient care. J Grad Med Educ. 2020;12(6):745–52. https://doi.org/10.4300/JGME-D-20-00030.1. Epub 2020 Dec 4. PMID: 33391599; PMCID: PMC7771597
10. Haidet P, Hempel EV, Louw BC, Chisty A. Virtual decisions: using a telehealth OSCE to enhance learners' triage skills. Med Educ. 2021;55(5):659. https://doi.org/10.1111/medu.14493. Epub 2021 Mar 17. PMID: 33733539; PMCID: PMC8250726.
11. Sartori DJ, Lakdawala V, Levitt HB, Sherwin JA, Testa PA, Zabar SR. Standardizing quality of virtual urgent care: using standardized patients in a unique experiential onboarding program. MedEdPORTAL. 2022;18:11244. https://doi.org/10.15766/mep_2374-8265.11244. PMID: 35497680; PMCID: PMC9001763
12. Lawrence K, Hanley K, Adams J, Sartori DJ, Greene R, Zabar S. Building telemedicine capacity for learners during the novel coronavirus outbreak: a case study and lessons learned. J Gen Intern Med. 2020;35(9):2675–9. https://doi.org/10.1007/s11606-020-05979-9. Epub 2020 Jul 8. PMID: 32642929; PMCID: PMC7343380
13. Sartori DJ, Hayes RW, Horlick M, Adams JG, Zabar SR. The TeleHealth OSCE: preparing learners to use telemedicine as a tool for transitions of care. J Grad Med Educ. 2020;12(6):764–8. https://doi.org/10.4300/JGME-D-20-00039.1. Epub 2020 Dec 2. PMID: 33391602; PMCID: PMC7771608
14. Boardman D, Wilhite JA, Adams J, Sartori D, Greene R, Hanley K, Zabar S. Telemedicine training in the COVID era: revamping a routine OSCE to prepare medicine residents for virtual care. J Med Educat Curri Develop. 2021;8:23821205211024076. https://doi.org/10.1177/23821205211024076. PMID: 34189270; PMCID: PMC8212360
15. American College of Physicians. Practice resources—telehealth guidance and resources. Philadelphia, PA; 2021. Accessed 30 Nov 2021. https://www.acponline.org/practice-resources/business-resources/telehealth
16. Noronha C, Lo MC, Nikiforova T, Jones D, Nandiwada DR, Leung TI, Smith JE, Lee WW. Society of General Internal Medicine (SGIM) education committee. Telehealth competencies in medical education: new frontiers in faculty development and learner assessments. J Gen Intern Med. 2022;37(12):3168–73. https://doi.org/10.1007/s11606-022-07564-8. Epub 2022 Apr 26. PMID: 35474505; PMCID: PMC9040701
17. Rutledge CM, O'Rourke J, Mason AM, Chike-Harris K, Behnke L, Melhado L, Downes L, Gustin T. Telehealth competencies for nursing education and practice: the four P's of telehealth. Nurse Educ. 2021;46(5):300–5. https://doi.org/10.1097/NNE.0000000000000988. PMID: 33481494; PMCID: PMC8395962
18. HHS. Guidance on nondiscrimination in telehealth: federal protections to ensure accessibility to people with disabilities and limited english proficient persons. Accessed Aug 2022. https://www.hhs.gov/civil-rights/forindividuals/disability/guidance-on-nondiscrimination-in-telehealth/index.html.

19. HHS. What should I know before my telehealth visit? telehealth.hhs.gov. Accessed 31 Mar 2024. https://telehealth.hhs.gov/patients/what-should-i-know-before-my-telehealth-visit.
20. Prasad A, Brewster R, Rajasekaran D, Rajasekaran K. Preparing for telemedicine visits: guidelines and setup. Front Med (Lausanne). 2020;7:600794. https://doi.org/10.3389/fmed.2020.600794. PMID: 33324665; PMCID: PMC7724018
21. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med. 2022;5(1):119. https://doi.org/10.1038/s41746-022-00663-0. PMID: 35982146; PMCID: PMC9387425
22. Ellaway RH, Coral J, Topps D, Topps M. Exploring digital professionalism. Med Teach. 2015;37(9):844–9. https://doi.org/10.3109/0142159X.2015.1044956. Epub 2015 Jun 1. PMID: 26030375
23. Lawrence K. Digital empathy in the age of coronavirus. The BMJ opinion. Accessed Aug 2022. https://blogs.bmj.com/bmj/2020/08/24/katharine-lawrence-digital-empathy-in-the-age-of-coronavirus/
24. Lyles CR, Wachter RM, Sarkar U. Focusing on digital health equity. JAMA. 2021;326(18):1795–6. https://doi.org/10.1001/jama.2021.18459.

10 Supporting Transitions Across the Medical Training Continuum with Simulation- and Performance-Based Assessment

Kinga L. Eliasz, Adina Kalet, Tavinder K. Ark, Elizabeth Wargo, Donna Phillips, Jeffrey Manko, and Sondra Zabar

Simulation- and Performance-Based Assessment to Support Learner Transitions

If not done with attention and care, the period of time when physicians transition to a new level of training can imperil patients, residency programs, healthcare systems, and individual clinicians. One of the biggest challenges for learners making training-level transitions is the need to adapt to new roles, responsibilities, and expectations. In the medical field, this often means taking on greater levels of autonomy, decision-making, and accountability. Medical professionals must also learn to navigate novel and complex systems, work with new colleagues, and communicate effectively with patients, families, and colleagues. Transitions can bring up feelings of uncertainty, anxiety, and self-doubt even for those trainees who had previously felt confident in their abilities. These feelings can be particularly acute and promote high-stakes consequences during the transition from undergraduate medical education (UME) to graduate medical education (GME, also referred to as "residency" or "internship"), as trainees move from being a student to being a physician in clinical practice settings, both in the hospital and clinical environment [1].

To help ease the transition and ensure readiness for clinical practice, New York University Grossman School of Medicine has created two immersive simulation events: Night-onCall (NOC) and First Night-onCall (FNOC). These simulations provide trainees with standardized yet realistic experiences in a safe and supportive environment, allowing them to make errors, receive feedback, and gain confidence as they prepare for their transition into residency. Through participating in deliberately designed simulated experiences, trainees confirm that they have the ability to make efficient and accurate clinical decisions under high-pressure situations.

Simulation is not only a valuable tool for easing learner transition from UME to GME, but it also plays a critical role in reducing the risk of harm to patients [2, 3], assessing clinical competence [4], and promoting a culture of safety within healthcare institutions [5]. Performance-based assessment (PBA) is a tool that can be leveraged to support transitions. PBA provides actionable data on individual student performance, team performance, and program gaps and strengths, and allows institutions to identify areas for improvement and customize training programs to meet the specific needs of trainees. By adjusting PBA to focus on individual competency-based assessments (in the case of NOC) or team/group-based assessments (with FNOC), trainees and teams receive tailored feedback and coaching, and the simulation experiences and related curriculum can be continuously improved over time.

This chapter provides a framework for designing and implementing simulations to support readiness-for-transitions and prepare for safe clinical practice. Below, we describe how PBA can be adjusted according to the assessment focus and specific needs at two inflection points during the transition from UME to GME: transitioning from

K. L. Eliasz (✉) · E. Wargo · S. Zabar
Department of Medicine, Division of General Internal Medicine and Clinical Innovation, New York University Grossman School of Medicine, New York, NY, USA
e-mail: sondra.zabar@nyulangone.org

A. Kalet
Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA

T. K. Ark
Data Science Institute, Medical College of Wisconsin, Milwaukee, WI, USA

D. Phillips
New York Simulation Center for Health Sciences, New York University Grossman School of Medicine, New York, NY, USA

J. Manko
Department of Emergency Medicine, New York University Grossman School of Medicine, New York, NY, USA

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_10

student to doctor just prior to medical school graduation and practicing collaborative healthcare teamwork immediately prior to beginning residency. With the help of simulations and PBA, trainees can be better prepared to take on their new roles, ultimately leading to better outcomes for patients, institutions, and professionals.

Two Unique Programs That Address Different Components of the UME to GME Transition

The role of medical educators in facilitating a successful UME–GME transition cannot be overstated. These educators play a critical role in designing and implementing programs that prepare trainees for the unique challenges of GME. Educators must be willing to explore and implement innovative approaches including leveraging technology, simulation, and data-driven feedback (e.g., PBA) to equip trainees with the skills and knowledge necessary to succeed in the complex and demanding world of GME.

A team including physicians, nurses, medical librarians, patient safety experts, and medical educators with experience in instructional design, simulation, and assessment developed two immersive simulation experiences—Night-onCall (NOC) [4] and First Night-onCall (FNOC) [5]. These events were deliberately designed and strategically implemented within the curriculum so that near-graduating and new interns had just-in-time [6] opportunities to rehearse their readiness for UME to GME transition in different contexts. The situations and instructional strategies utilized in NOC and FNOC differ (Table 10.1); NOC emphasizes demonstrating individual clinical competence, while FNOC focuses on collective competence demonstrated in how they approach team work and patient safety.

Night-onCall (NOC): UME Readiness-for-Internship Simulation Event [4]

In 2014, the Association of American Medical Colleges (AAMC), responsible for accrediting medical schools in the United States, released a set of 13 core Entrustable Professional Activities (EPAs)—behaviors that all entering residents should be expected to perform on day 1 of residency without direct supervision, regardless of specialty choice. They provided detailed guidance meant to drive the UME community toward refining, measuring, and benchmarking the minimal level of competence in each of these domains expected of a medical school graduate [7, 8]. NYU Grossman School of Medicine is one of the ten pilot schools in the country investigating EPAs in the UME context. Using the 13 core EPAs guidelines, we built and studied an immersive simulated "on call" experience for near-graduate medical students to measure readiness-for-internship, called Night-onCall (NOC). NOC is currently integrated into the curriculum prior to graduation. The goal of NOC has been to provide near-graduating medical students with an *individual* immersive simulation experience that both assesses their performance related to the 13 core EPAs and supports their learning as they prepare to transition to residency. An *individual* experience was chosen for two fundamental reasons: (1) assessment (i.e., measuring individual competence) and (2) providing learners with an opportunity to experience the decisional autonomy of an intern on call, an experience that is rare during UME training.

NOC is a multi-station OSCE. NOC aligns with the literature that supports the utility of a well-designed OSCE as a valid assessment of clinical competence, assuming careful attention is paid to "contextual fidelity," including the interprofessional nature of most medical work and accurate "professional role reproduction" [9]. During what was originally a 4-hour experience, learners see four SP cases with varying degrees of complexity, each of which requires first answering a call from a standardized nurse (SN), then evaluating an SP with the SN in the room and making immediate management decisions, and lastly, writing a coverage note. In the original version of NOC, between the first and second patient case, the learner engaged in a triple-jump exercise [10] by completing a content-specific e-learning module [6]. The first patient case also requires that the learner makes a phone call to a standardized attending (SA, portrayed by an experienced clinician) to orally present and discuss the case. The NOC experience ends with the learner completing a handoff of the four patient cases to a standardized intern (SI, portrayed by a senior medical student). NOC has been modified to a 3-hour version with 3 SP cases to facilitate large graduating classes without compromising measure of the 13 EPAs. This has made NOC logistically more feasible by assigning the e-learning module as pre-work and reducing the number of clinical cases to three. Funded by the Josiah Macy Foundation, we have built a NOC consortium of ten US medical schools. The goal of the consortium is to collaborate and establish a benchmark for near-graduate learners. We have collected data on 1620 consented learners over 7 years (as of October 2024).

Throughout the NOC experience, learners receive competency and entrustment assessments from multiple perspectives including standardized patients, standardized nurses, standardized attendings, and standardized interns, all of whom are rigorously calibrated to a standardized simulation activity. A total of 14 assessments, based on tools with validated evidence where available [4], are used during NOC to score behavioral competencies in real time. A confirmatory factor analysis provided early evidence that entrustment judgments can be made based on a student's NOC performance [11].

Table 10.1 Similarities and differences between Night-onCall (NOC) and First Night-onCall (FNOC)

	Night-onCall (NOC)	First Night-onCall (FNOC)
Event descriptor	An immersive simulation with activities structured to assess and address MD competencies (i.e., 13 core Entrustable Professional Activities (EPAs))	An institutional patient safety event with activities structured to replicate and assess key patient safety skills
Curriculum point	Undergraduate Medical Education (UME), few months prior to graduation	Graduate Medical Education (GME), during intern orientation
Target learner	Near-graduating medical students	Incoming interns
Event framework	Objective Structured Clinical Exam (OSCE)	Group Objective Structured Clinical Exam (GOSCE)
Event length	4- or 3-hour individualized versions (3 learners, 6SPs, 1 attending/OSCE rotation)	4- or 3-hour group versions plus preassigned e-learning modules (up to 40 learners/event)
Event goal	Individualized experience focused on *individual* competence (developing an individual mental model)	Team experience focused on *collective* competence (developing a shared mental model)
	Focused on providing learners with autonomy to be an intern on call	Focused on creating a culture of safety by introducing and practicing escalation protocols, emphasizing principles of high reliability and national patient safety goals, and networking with new colleagues
Event activities (in chronological order)	4 standardized patient cases (**oliguria 1, oliguria 2, hypertension, informed consent) that target case-specific skills such as history gathering, physical exam, patient care, and EPA behaviors	3 standardized patient cases (hypotension, *informed consent, fever evaluation), the escalation call is embedded into the hypotension case and the medical error is embedded into the fever evaluation with incorrect patient labels
	Coverage note documentation following each case	Coverage note documentation after the informed consent case
	Oral presentation to a standardized attending (via phone call)	2 acute manikin cases (hypoglycemia and anaphylaxes) with debriefing or PPE donning and doffing and treatment of anaphylaxis
	Evidence-based medicine activity (literature search)	Patient handoff (IPASS) of the 2 cases and GOSCE debrief
	Culture of safety activity (paper scenario)	*Patient safety room (recognize common patient safety hazards in a simulated patient room)
	Patient handoff of the 3 or 4 cases	*Patient safety rounds case-based discussion
	NOC debrief (learner also receives feedback report and completes a learning plan activity)	
E-learning module integration	**WISE-onCall module for oliguria is embedded into the NOC experience as a just-in-time learning strategy to facilitate the "triple-jump' approach: (1) Learner completes the first oliguria case in a discovery learning fashion (2) Learner completes the WISE-onCall oliguria module, which provides just-in-time learning around oliguria and is meant to provide an overall framework to approach patient cases (3) Learner completes a second oliguria case (different case) to solidify the learning from steps 1 and 2	WISE-onCall modules are assigned prior to the FNOC experience as a preparation for future learning strategy [21]—providing the learner with an overall framework to approach patient cases (different case examples) and establishing a baseline and expectations for incoming interns
Assessment and feedback	Every individual learner's performance is independently measured on the 13 core EPAs by multiple standardized raters including patients, nurses, attendings, and interns	GOSCE team performance is measured by standardized patients and nurses
	Every learner receives a comprehensive individualized feedback report immediately following the NOC experience	Faculty facilitators provide groups with specific feedback and all learners receive a resource guide with information tied to cases and activities
Reflection phase	After the learner receives feedback report, reflection is embedded in the individualized learning plan activity	Reflection exercise follows every FNOC activity (learner is asked to reflect on what was most memorable about each given activity and why)

**Eliminated in COVID 3-hour Version*
***Eliminated when shortened NOC to 3 hours*

Immediately following the NOC experience, each learner receives an individualized report (Fig. 10.1), which provides detailed feedback on their performance and is designed to highlight patterns of competence across activities. Learners also participate in a debriefing session with faculty, during which they develop their own individualized learning plans to reinforce strengths and address identified weaknesses prior to beginning residency. During this session, each learner is also provided with a resource guide, which includes voluntary mentorship opportunities. We designed the debrief session and resources in this manner, in order to empower individual learning and offer learners a supportive environment to address tailored, actionable goals. Medical schools received aggregate reports of their class EPA performance for quality assurance and guidance for annual curricular changes.

Beyond the goal of assessing individual competence, NOC provides near-graduating medical students with a lived experience of the expectations associated with their new level of autonomy in a psychologically-safe and supportive learning space. By acting as the intern in NOC, medical students gain firsthand experience in responsibility, accountability, interacting with different members of the healthcare team, managing multiple patients, and delivering information in a concise, clinically-meaningful manner while also tailoring it to different team members (i.e., nurses, patients, supervisors, colleagues). NOC is an invaluable experiential tool for preparing near-graduating medical students for the challenges of residency and ensures that each trainee is equipped with the clinical skills and knowledge they need to function effectively in an increasingly independent role.

First Night-onCall (FNOC): GME Patient Safety Orientation Simulation Event [5]

Readiness to take care of patients at the beginning of residency varies greatly based on which medical school an intern graduated from. As a result, program directors have often implemented simulation-based transition programs, or "bootcamps," to ensure these novice practitioners are uniformly prepared [12–14]. However, residency requires collaboration in addition to individual competence to ensure a culture of safety. To facilitate the transition from UME to GME, NYU Grossman School of Medicine and the New York Simulation Center for the Health Sciences (NYSIM) developed FNOC for all incoming interns. FNOC is a collaborative, immersive "on call" simulation experience that focuses on actions that enhance patient safety and introduces the principles, behaviors, and tools that NYU Langone Health has adopted to promote safety, quality, and patient experience: escalation, recognizing and reporting a medical error, and creating a culture of safety and continuous learning from day one.

We chose a multi-station Group Objective Structured Clinical Exam (GOSCE) simulation experience for FNOC in order to provide all incoming interns from different medical schools the opportunity to experience developing *collective* competence and skills. Collective or distributive competence is the capacity of teams or groups, that goes beyond accumulated individual competence of group members, and requires engagement and exquisite verbal communication among individuals to ensure a culture of safety [15]. The underpinning of this approach to competence is defined in situated learning theory [16], socio-material [17], and systems theory [18, 19] and suggests that competence emerges through shared experience. In this case, simulation is used to provide trainees with an opportunity to become more comfortable with their newly-acquired professional roles (i.e., interns) alongside their peers and other healthcare professionals, in a supportive learning environment. FNOC also creates a culture of safety by introducing and practicing escalation protocols (such as those needed for decompensating patients requiring rapid response teams), emphasizing high reliability care and the importance of national patient safety goals (e.g., two patient identifiers, the need for common procedures such as structured handoffs), and networking with new colleagues.

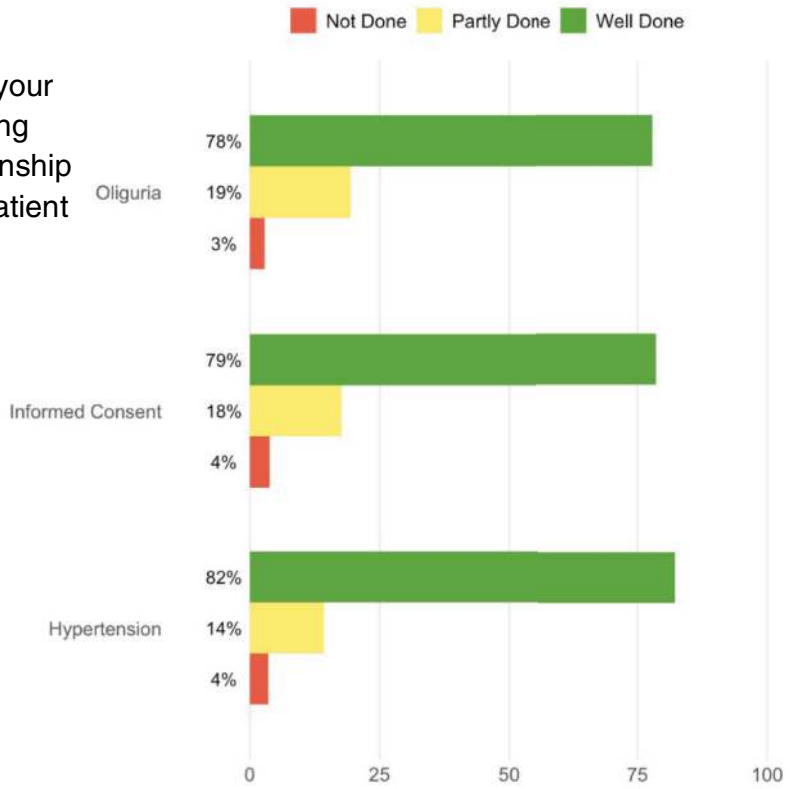
Prior to the FNOC event, incoming interns are assigned a set of WISE-onCall online modules as part of the onboarding process, which provide them with a framework for approaching acute inpatient complaints ([6]; https://med.nyu.edu/education/other-educational-programs/wise-program). The purpose of assigning modules prior to FNOC is both to prepare interns to learn from the simulation by providing frameworks for addressing acute clinical issues, therefore consolidating prior knowledge [21], and to communicate institutional expectations for quality and safety.

During the 4-hour FNOC experience, new interns, in groups of four, participate in a GOSCE to conduct an informed consent conversation, evaluate a decompensating patient where they need to activate a rapid response team (escalation), document a clinical encounter, face an embedded medical error, and end with a handoff of a patient to an incoming team. Following the simulation, they participate in a simulated, case-based discussion on patient safety rounds with institutional patient safety leaders.

In 2020, we redesigned the program to address topics relevant to the COVID-19 pandemic with new content and by changing the design to protect learners from infection (e.g., smaller sized learner groups). A manikin-based simulation case was added and included assessing a patient and donning and doffing personal protective equipment (PPE). The case-based discussion and informed consent GOSCE case were eliminated, and environmental hazards curricula were modified. Education and safety leaders recognized FNOC as an

Communication

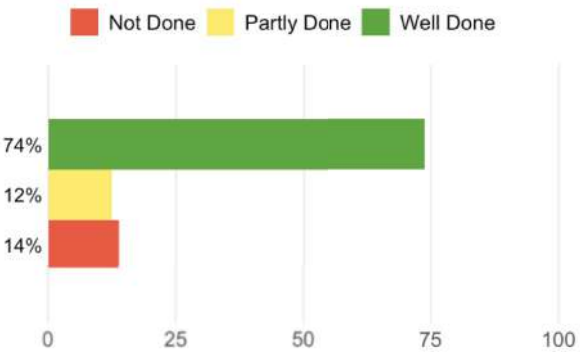
Standardized patients rated your communication skills, including information gathering, relationship building, organization, and patient education and counseling.



Patient Handover

A standardized intern rated your handoff for the four patient cases according to the IPASS framework

I - Illness Severity
P - Patient Summary
A - Action List
S – Situation Awareness/Contingency Planning
S - Synthesis by Receiver



Entrustment

The standardized intern also indicated their assessment of the quality of your situational overview, an OVERALL assessment of whether you transferred incorrect information, omitted important information, or engaged in tangential conversation and the level of trust (Entrustment) based on the information you delivered.

Strong *I would trust this intern with patient care even without supervision*

Fig. 10.1 Score card report delivered immediately following Night-onCall (NOC)

opportunity to address new public health concerns and additional curricula focus such as equity and patient experience as part of safe, high-quality patient care. The current version of FNOC is 3 hours long. Learners still work as a team to interact with standardized nurses and standardized patients, who assess them as a group.

The GOSCE during the FNOC experience is focused on collective competence, and each team receives specific feedback from multiple perspectives including the standardized patients, nurses, and faculty facilitators. After each activity, learners are asked to identify in writing what was most memorable about that activity, so that they have an opportunity to pause and reflect on the specific learnings and takeaways from that activity. FNOC ends with a faculty-led debriefing session where interns are encouraged to voice their feelings about their transition, ask clarifying questions, and identify action items they plan to incorporate on day one of internship. Following the FNOC event, all incoming interns are provided an electronic resource guide with institution-specific protocols, support team details and contact information, so that they feel empowered and supported as new members of the institution. This is also available to them on the institution's learning management platform.

The NYU Grossman School of Medicine has hosted FNOC for 8 years and delivered this experiential patient safety orientation to all incoming interns. We provide the safety orientation to 200+ interns a year (1600+ total learners over 8 years) and now are assured that all residents at our institution have had a common patient safety orientation and practice.

Over time, we have seen a cultural change at our institution: program coordinators contact FNOC leadership to plan their program orientation and bootcamp around FNOC, core faculty and program directors all expect to participate as faculty observers, and patient safety leaders suggest relevant modifications to our resource guide. FNOC has also contributed to faculty scholarship and recognition and receives ongoing support from the academic health center. In all these ways, an annual intern experiential patient safety curriculum has exponential benefits to the healthcare community.

Conclusions

The use of simulation and PBA can greatly aid trainees in their transition from UME to GME. These tools allow trainees to practice and refine their skills in a psychologically-safe and controlled environment while also receiving immediate feedback from rigorously-calibrated raters. Incorporating simulation and PBA into medical education is a vital step toward producing competent, safe, and confident physicians.

The opportunity to rehearse authentic professional activities and new role responsibilities during NOC and FNOC, each focused on different but complementary aspects of readiness-for-internship, helps trainees make a smooth transition into residency. We made instructional design decisions informed by the unique challenges of both UME and GME.

Simulation can be used reliably and effectively to smooth the UME to GME transition, providing learners a safe learning space to "put it all together." We intentionally designed this as a series of simulations to ensure incoming interns recognize their role on the healthcare team to provide safe, high-quality care for patients. Simulation helped our team provide transitioning trainees with not only standardized and supportive educational experiences but also data-driven feedback to guide their personal preparation for GME. With support from the Josiah Macy Foundation, we are building a consortium of medical schools that are working together to further refine and develop the use of immersive simulations like NOC and FNOC with goals to collaborate on building a robust, adaptable, and feasible transition curriculum while also providing assessment, benchmarking, and standard-setting to establish a validity argument for using the data to assess readiness for transitions (for more information, see https://www.nightoncall.org). In these ways, we introduce our trainees, both individually and collectively, to their role as good citizens of a safe, high-quality healthcare system.

References

1. Chang LY, Eliasz KL, Cacciatore DT, Winkel AF. The transition from medical student to resident: a qualitative study of new residents' perspectives. Acad Med. 2020;95(9):1421–7. https://doi.org/10.1097/ACM.0000000000003474.
2. Barsuk JH, Cohen ER, Wayne DB, Siddall VJ, McGaghie WC. Developing a simulation-based mastery learning curriculum: lessons from 11 years of advanced cardiac life support [published correction appears in Simul Healthc. 2016 Apr;11(2):147]. Simul Healthc. 2016;11(1):52–9. https://doi.org/10.1097/SIH.0000000000000120.
3. Cohen ER, Feinglass J, Barsuk JH, Barnard C, O'Donnell A, McGaghie WC, Wayne DB. Cost savings from reduced catheter-related bloodstream infection after simulation-based education for residents in a medical intensive care unit. Simul Healthc. 2010;5(2):98–102. https://doi.org/10.1097/SIH.0b013e3181bc8304.
4. Kalet A, Zabar S, Szyld D, et al. A simulated "Night-onCall" to assess and address the readiness-for-internship of transitioning medical students. Adv Simul (Lond). 2017;2:13. Published 2017 Aug 14. https://doi.org/10.1186/s41077-017-0046-1.
5. Eliasz KL, Kalet A, Buckvar-Keltz L, et al. A culture of safety from day 1: an institutional patient safety initiative to support incoming interns. J Grad Med Educ. 2018a;10(3):347–9. https://doi.org/10.4300/JGME-D-17-00828.16.
6. Szyld D, Uquillas K, Green BR, et al. Improving the clinical skills performance of graduating medical students using "WISE OnCall," a multimedia educational module. Simul Healthc. 2017;12(6):385–92. https://doi.org/10.1097/SIH.0000000000000254.
7. Association of American Medical Colleges (AAMC). Core Entrustable Professional Activities for Entering Residency,

Curriculum Developers Guide. 2014. https://store.aamc.org/downloadable/download/sample/sample_id/63/. Accessed 2 Jan 2022.
8. Englander R, Flynn T, Call S, et al. Toward defining the foundation of the MD degree: core entrustable professional activities for entering residency. Acad Med. 2016;91(10):1352–8. https://doi.org/10.1097/ACM.0000000000001204.
9. Hodges B. Validity and the OSCE. Med Teach. 2003;25(3):250–4. https://doi.org/10.1080/0142159031000100283 6.
10. Smith RM. The triple-jump examination as an assessment tool in the problem-based medical curriculum at the University of Hawaii. Acad Med. 1993;68(5):366–72. https://doi.org/10.1097/00001888-199305000-00020.
11. Eliasz KL, Ark TK, Nick MW, Ng GM, Zabar S, Kalet AL. Capturing entrustment: using an end-of-training simulated workplace to assess the entrustment of near-graduating medical students from multiple perspectives. Med Sci Educ. 2018b;28(4):739–47. https://doi.org/10.1007/s40670-018-0628-0.
12. Frischknecht AC, Boehler ML, Schwind CJ, et al. How prepared are your interns to take calls? Results of a multi-institutional study of simulated pages to prepare medical students for surgery internship. Am J Surg. 2014;208(2):307–15. https://doi.org/10.1016/j.amjsurg.2014.01.014.
13. Goren EN, Leizman DS, La Rochelle J, Kogan JR. Overnight hospital experiences for medical students: results of the 2014 clerkship directors in internal medicine national survey. J Gen Intern Med. 2015;30(9):1245–50. https://doi.org/10.1007/s11606-015-3405-4.
14. Schwind CJ, Boehler ML, Markwell SJ, Williams RG, Brenner MJ. Use of simulated pages to prepare medical students for internship and improve patient safety. Acad Med. 2011;86(1):77–84. https://doi.org/10.1097/ACM.0b013e3181ff9893.
15. Lingard L. Rethinking competence in the context of teamwork. In: Hodges B, Lingard L, editors. The question of competence: reconsidering medical education in the twenty-first century. Ithaca: Cornell University Press; 2012. p. 42–69.
16. Lave J, Wenger E. Situated learning: legitimate peripheral participation. Cambridge University Press; 1991.
17. Fenwick T, Dahlgren MA. Towards socio-material approaches in simulation-based education: lessons from complexity theory. Med Educ. 2015;49(4):359–67. https://doi.org/10.1111/medu.12638.
18. Boreham N. A theory of collective competence: challenging the neo-liberal individualisation of performance at work. Br J Educ Stud. 2004;52(1):5–17. https://doi.org/10.1111/j.1467-8527.2004.00251.x.
19. Engeström Y, Middleton D. Cognition and communication at work. Cambridge: Cambridge University Press; 1998.
20. https://med.nyu.edu/education/other-educational-programs/wise-program
21. Mylopoulos M, Brydges R, Woods NN, Manzone J, Schwartz DL. Preparation for future learning: a missing competency in health professions education? Med Educ. 2016;50(1):115–23. https://doi.org/10.1111/medu.12893.
22. https://www.nightoncall.org/

Objective Structured Teaching Exercises (OSTEs) from the Teacher's Perspective: What, Why, When, and How?

11

Cynthia Osman and Anne Dembitzer

What Is an OSTE?

An Objective Structured Teaching Exercise (OSTE) is a variation on the traditional OSCE, where participants practice and/or demonstrate their teaching skills. In an OSTE, actors play standardized learners (SLs), rather than (or in addition to) SPs. OSTEs aimed at providing practice and formative assessment of teaching begin with a brief didactic orientation session to prepare participants for the scenarios. Next, participants enter a room with an SL and engage in a teaching scenario. Immediately following the scenario, participants reflect on their own performance, and the SL provides feedback to the participant. Sometimes, faculty or peer observers share their feedback as well. Often a group debriefing follows the scenario(s). Like OSCEs, OSTEs can also be videotaped for later review, providing additional opportunities for reflection and self-assessment.

In Group OSTEs (GOSTEs), participants engage in the simulation in teams. GOSTEs have several benefits: participants have the opportunity to share and discuss strategies prior to beginning the scenarios, they can observe and learn from their peers, fewer SLs are needed per participant, and scenarios simulate team conversations typical in health care.

Why Use OSTEs?

In 2010, the Carnegie Foundation for the Advancement of Teaching called for increased integration of formal learning with clinical experience, highlighting the importance of clinician educators to the educational mission of academic medical centers [1]. It is widely recognized that sophisticated faculty development is needed to ensure that clinician educators are able to respond to the increasingly complex teaching expectations [2].

Institutions are expanding programming in faculty development [3–5]. While there are many reasons for this, a significant factor are mandates from the Liaison Committee on Medical Education ([6], Standard 4.5) and the Accreditation Council for Graduate Medical Education [7, 8] that state that residency and fellowship faculty development programs address teaching skills. In addition, the number of resident-as-teacher programs have grown in the last 20 years [9], likely due to similar mandates for resident-as-teacher programming from the LCME ([6], Standard 9.1) and ACGME (core requirement IV.B.1.e).(1).(d), 2021), which states that "Residents must demonstrate competence in . . . educating . . . students, residents and other health professionals."

Despite this growth in programming around formal teaching skills, few clinical educators at any level have had their teaching skills assessed in authentic settings in real time [4, 5, 9]. As Boillat et al. [10] states, ". . .the OSTE incorporates many of the elements of an 'ideal' faculty development tool by: (1) creating authentic teaching contexts; (2) enhancing the objective assessment of teaching skills with SLs and predetermined criteria; (3) allowing direct feedback from learners and peers; and (4) providing an opportunity for repeated practice."

Additionally, OSTEs are an effective means for faculty to simultaneously learn new core curricular content *and* practice their teaching skills. For example, there is mounting evidence that faculty and residents lack the skills needed to discuss the multiple forms of bias in health and the healthcare workplace [11], despite the increasing availability of relevant curricula [12–16] As Acosta and colleagues have noted, "both cultural sensitivity training and unconscious bias training are important, but faculty need more. HP [health professions] faculty need not only the ability to recognize prejudice and discrimination but also the tools to speak up against it when they witness it [11]." OSTEs can provide this

C. Osman (✉)
Department of Pediatrics, New York University Grossman School of Medicine, New York, NY, USA
e-mail: cynthia.osman@nyulangone.org

A. Dembitzer
Department of Medicine, Division of General Internal Medicine and Clinical Innovations, New York University Grossman School of Medicine, New York, NY, USA

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_11

psychologically safer learning environment compared to the clinical setting.

When Should Medical Educators Consider Using an OSTE?

We have found OSTEs to be particularly valuable to:

1. Identify teaching skill gaps
2. Learn teaching skills
3. Assess teaching skills
4. Conduct program evaluation of teaching programs

Identify Teaching Skill Gaps

The Education for Educators (E4E) program, a longitudinal faculty development program (FDP) in medical education at New York University Grossman School of Medicine (NYUGSOM), has used OSTEs and GOSTEs to facilitate learners' identification of their own teaching gaps in order to prepare them to gain the most from educational sessions. By writing personal learning goals at the conclusion of the OSTE, E4E faculty participants demonstrate their ability to recognize their personal teaching skill gaps [17]. Trowbridge's systematic review of OSTEs [18] and Stone and colleagues in their work describing the use of OSTEs in a faculty development program [19] have similarly noted that faculty who participate in OSTEs report greater recognition of teaching strengths and gaps.

Learn Teaching Skills

Stone et al. [19] found that faculty reported believing they were able to improve their skills and develop new teaching strategies as a result of participating in OSTEs. OSTEs provide a psychologically safe setting for active learning experiences, enhance motivation to learn, and promote deliberative practice [20, 21]. Participants have an opportunity to try new skills, receive feedback, and try again. The standardized structure of OSTEs assures that all participants receive a similar learning experience facilitating a shared understanding of excellent teaching skills and creating a culture that supports continued learning.

Assess Teaching Skills

OSTEs provide opportunities to assess participants' teaching skills. Performance is often evaluated by SLs and/or by observing faculty or peers. Evaluation is usually done using a rubric that assesses a number of teaching skills or behaviors (see Appendix R). OSTEs are performance-based assessments of skills, considered a step closer to actual practice than a knowledge-based assessment would be, and are therefore predictive of actual practice or "workplace-based" implementation, as illustrated in Miller's Triangle of Learner Assessment strategies [22].

Conduct Program Evaluation of Teaching Programs

Data from OSTEs, if collected with attention to measurement quality, may provide evidence supporting the impact of educational programs at the midrange levels of learning and/or behavior in the Kirkpatrick Model of program assessment [23].

How Have Programs Used OSTEs to Teach or Assess Clinician Educators' Skills?

Table 11.1 details some examples of programs utilizing OSTEs to teach and/or assess a variety of teaching skills.

How Do You Optimize Faculty and Resident Participation in an OSTE?

The instructional design of an OSTE is uniquely well-suited for experienced adult learners who are highly self-directed and appreciate experiential learning of practical skills that are immediately applicable. However, there are some challenges to engaging faculty and residents in OSTE-based faculty development experiences. These are detailed along with potential solutions in Table 11.2.

What Are the Steps to Creating an OSTE Program?

The creation and implementation of an OSTE and an OSCE are fundamentally similar [10, 30], but we outlined a few specific considerations for ensuring effective and engaging OSTEs based on 20 years of collective experience:

1. *Clarify Goals of Creating an OSTE Program*: Creation and implementation of OSTEs are a large undertaking, requiring significant time to schedule, create cases and assessment tools, and train SLs [30]. Therefore, it is important to confirm that the OSTE format fits your goals.

2. *Obtain Support*: Support from leadership is key to successful OSTE implementation. Departmental leaders should be reminded that faculty development is mandated by the ACGME and LCME. Organizers should obtain both financial support (to pay actors for SLs) and logistical support (to free up resident, fellow, and/or faculty time for participation in the exercise) early in the planning process.

Table 11.1 Evidence from literature on OSTEs

Article title	OSTE goal and format	Evaluation	Findings
Faculty development on professionalism and medical ethics: The design, development, and implementation of Objective Structured Teaching Exercises (OSTEs) [24]	Lu and colleagues created a faculty development program that included OSTEs to teach faculty how to address issues of professionalism and medical ethics with medical students.	Participants completed a post-workshop satisfaction survey and a pre−/post-survey on confidence and attitudes about teaching professionalism and medical ethics.	As a result of participation, faculty ($n = 20$) reported greater confidence in several areas including their ability to communicate concerns and handle professional issues that involve medical students.
The impact of an objective structured teaching evaluation on faculty teaching skills [25]	Julian and colleagues conducted a faculty development program using a series of OSTEs to teach skills such as providing feedback to a learners, bedside teaching, and outpatient precepting.	Participants completed a 21-item pre−/post-survey on 7 general teaching skills domains: learning climate, control of session, communication of goals, promotion of understanding and retention, feedback, evaluation, and promotion of self-directed learning [26].	Faculty ($n = 46$) reported significant improvement of their teaching skills in all 7 areas. However, student evaluations of the participating faculty's teaching prior to the OSTE program and 6 months after completion did not change.
Using the objective structured teaching ecounter to assess resident teaching skills [27]	Oh and colleagues utilized OSTEs to assess family medicine resident's ability to teach medical students using the 5 microskills of precepting. Residents received instruction on use of the 5 microskills and then participated in an OSTE.	SLs rated residents' use of the 5 microskills. Authors hosted focus group to solicit feedback from the residents.	Residents ($n = 41$) were able to demonstrate use of the 5 microskills. Residents also reported greater teaching confidence after participation in the OSTE.
Creating an Objective Structured Teaching Examination to Evaluate a Dental Faculty Development Program [28]	McAndrew and colleagues developed and implemented a 3-station OSTE to assess the impact of participation in a dental faculty development program that focused on conflict resolution, peer teaching, and small group facilitation.	SLs completed behaviorally anchored checklists on case-specific items pre- and post-curriculum implementation.	Faculty members' ($n = 12$) general teaching skills and case-specific teaching skills improved after the faculty development program.
A pilot randomized, controlled trial of a longitudinal residents-as-teachers curriculum [29]	Morrison and colleagues assessed the impact of a resident-as-teacher curriculum using pre- and post-program OSTEs. Resident were randomly assigned to control group or to participate in the resident-as-teacher program; both groups completed the OSTEs.	SLs evaluated resident performance during the OSTEs on 5-point Likert scale.	Residents ($n = 23$) who participated in the resident-as-teacher curriculum demonstrated a 22.3% improvement in their skills on the post-curriculum OSTE. They also performed better than residents who did not complete the teaching program.
Development and implementation of an objective structured teaching exercise (OSTE) to evaluate improvement in feedback skills following a faculty development workshop [19]	Stone and colleagues employed OSTEs to assess the impact of participation in a faculty development workshop focused on feedback skills through OSTEs. Participants were randomly assigned to complete the OSTE before the workshop or after the workshop.	SLs completed a 13-item checklist on participants.	Authors did not find statistically significant differences in observed teaching performance between the two groups of faculty ($n = 56$). Possible reasons include the participants not having enough time to develop and execute the new skills or participants with prior educational training skewing the pre-intervention OSTE group.

3. *Consider Resources*: OSTE planning is similar to that of OSCEs (refer to the information on planning an OSCE timeline in Chap. 2, Step 2, and Table 2.5). Detailed consideration of the resources needed for completion of the OSTE is essential; this includes: space for the event (either in a simulation center or other venue), financial support, identification of observers (if needed), identification of facilitators for the introductory and/or debriefing sessions, and planning for training of the SLs.
4. *Perform Needs Assessment*: A needs assessment should include data collection such as surveys, interviews, and/or focus groups with key stakeholders. Stakeholders can provide input on teaching skill gaps and include medical school deans, program directors, division directors, chief residents, trainees, and teaching faculty. The needs assessment process can simultaneously serve to garner leadership support for and buy-in to the program. Among the competencies for medical educators that can serve as the

Table 11.2 Engaging faculty and residents in OSTEs

Learner	Challenge	Possible solution
Faculty	Faculty do not have provided protected time to attend faculty development OSTE sessions	Obtain support from departmental and/or divisional leadership, with a reminder that faculty development on teaching skills is a mandate of both the ACGME and the LCME.
	Differences between number of faculty that sign-up and actual participation	Offer multiple sessions at different times on the same topic, timing sessions at maximally convenient times (such as summer Grand Rounds time), and utilizing a GOSTE format rather than individual OSTEs.
	Depending on their age, temperament, and prior experience, faculty are not familiar nor comfortable with being observed by and receiving feedback from others	The GOSTE format is especially useful for faculty participants. Working in groups can decrease the inevitable stress of being "on the hot seat," and many faculty appreciate the rare opportunity to observe each other in action.
Resident	Residents perform different kinds of teaching than faculty	Tailor OSTE scenarios and teaching challenges to participants.
	Residents have complex schedules	Organizers should be sure to arrange sessions so most of the residents can participate, ideally during the program's curricular time. Most resident OSTEs are a mandatory experience, so attendance should be guaranteed within the confines of the residents' clinical schedules.

focus of an OSTE are medical (or content) knowledge, learner-centeredness, interpersonal and communication skills, professionalism and role modeling, practice-based reflection, and systems-based practice [31]. Teaching skills commonly addressed in OSTEs include giving feedback, precepting, professionalism, implicit bias, working with challenging learners, working with impaired learners, and working with excellent learners. As with OSCE development, it is recommended that the OSTE designers develop the program to ensure that the targeted teaching skills are addressed as well as additional relevant issues such as content or interpersonal communication skills (see sample OSCE blueprint: Chap. 2, Step 3, Table 2.6).

5. *Choose a Format: OSTE* Versus *GOSTE*: As a reminder, group OSTEs (or GOSTEs) allow participants to work in pairs or larger groups. Depending on the types of participants, the finances available, and the specific goals of the OSTE, a group OSTE format may be appropriate. GOSTEs have the advantage of requiring fewer actors per participant, enabling practice of team communication and allowing for greater flexibility in scheduling if the number of participants varies at different sessions.

   GOSTE formats may vary. Participants can work in pairs, where one participant "takes the lead" while the other participant acts as backup, reducing performance pressure. In the NYUGSOM E4E program, groups of three or four faculty participants rotate through a series of OSTE stations together. Participants review the OSTE scenario, plan the teaching approach as a group, and rotate taking the lead for each station. Feedback is shared by observing participants, standardized learners and patients, and an E4E faculty facilitator.

6. *Create Cases*: Case materials, which include SL instructions, SL checklists, and participant instructions, should be developed and reviewed for authenticity in consultation with stakeholders. The case should be piloted during the development process allowing for SLs to contribute to further modifications of the case materials. Clerkship directors, residency and fellowship program directors, and chief residents can be particularly helpful to flesh out the details of common teaching challenges for faculty. The more realistic and detailed the scenario, the more believable and useful it is for the participants (see Appendix S for example SL instructions).

   Like OSCE cases, OSTE case materials should including setting and learner profile including prior academic performance, experience in the clinical learning environment, and career aspirations. Clinical details if relevant, may be necessary, for instance, in a precepting scenario.

   Once the case has been fleshed out, the OSTE organizers must either choose or create a checklist for the SL to use when assessing the participant. Checklists are key in helping to standardize the feedback process and to ensure each participant receives feedback on the same core domains. There are sample checklists in the literature (see

Osman et al. [32]), or see our example of a generic OSTE checklist in Appendix R.

Finally, if participants will be observed by others besides the SL, then OSTE organizers must create guiding materials for the observers to assess the participants' performance.

7. *Selection of Standardized Learners*: OSTE developers should keep their goals in mind when considering whether to recruit actors without a healthcare background or healthcare trainees or professionals to serve in the SL role. There are tradeoffs. Those with relevant experience and a medical fund of knowledge might be expected to portray a medical student or resident more authentically. On the other hand, actors less familiar with and beholden to the medical hierarchy may feel freer to provide frank feedback to faculty or resident OSTE participants [33]. Additionally, actors are more expert at consistent case portrayal as well as being more comfortable with the inevitable improvisation required during an OSTE encounter.

   The choice depends on resources (actors are usually compensated, while healthcare trainees might be enticed to volunteer), type of teaching scenarios (a case with complex medical details might best be played by trainees or healthcare professionals), and need for reliable and valid performance ratings (actors are easier to train to use checklists of behaviorally specific items).

   If using health professions learners from local programs, consider recruiting from a different discipline than participants (e.g., nursing students for physician faculty), and be sure to inform and remind participants that the SL's are playing a role.

8. *Choose Sources of Feedback*: There are many possible sources of feedback for OSTE participants. In some OSTE scenarios, there may be both an SL and an SP. For example, if the OSTE scenario involves teaching at the bedside, both the SP and SL will need training. Faculty observers might provide clinically oriented feedback to participants. This observation can be conducted directly, through a one-way mirror, or based on review of a videotape of the interaction or sitting in the corner of the room with direct observation. Howell and colleagues have utilized "family faculty," parents of pediatric patients who employ their personal experiences in healthcare to provide feedback to pediatric residents about their interpersonal skills [34]. And, finally, participants can receive feedback from fellow participants if a GOSTE format is used.

9. *Train SLs and Observers*: As with training SPs for OSCEs (see Chap. 2, Step 6), training SLs should include case portrayal, checklist, and feedback. Case portrayal training typically includes a read-through of the case, with time for questions from the SLs, as well as a series of role-plays with the trainer who can vary the skill level represented.

   The training of SLs (and of observers and/or SPs) should include a thorough discussion of the checklist tool, with opportunities for SLs/SPs/observers to assess different levels of learners so that they can calibrate their ratings to levels of performance.

   Finally, the SLs (and observers and/or SPs, if they will be giving feedback to the participants) should be prepared to use the common feedback framework used at your institution. One such model is that from Gigante et al. [35] which includes the following five steps: (1) outline the expectations for the learner, (2) prepare the learner to receive feedback, (3) ask the learner for self-assessment, (4) tell the learner how he or she is doing, and (5) agree on a plan for improvement.

10. *Establish the Role of the OSTE Leaders*: Most OSTE sessions begin with an introduction to the goals of the session, the logistics of the cases, and any teaching content that might be helpful to the participants as they rotate through the cases. It is important to set ground rules for the discussion, such as keeping information about fellow participants' performance confidential. However, it is helpful to keep the introductory portion of the session brief since the more important part of the experience for the participants will be engaging with the SLs in the actual scenarios, receiving feedback, and then debriefing the experience.

    The debriefing discussion is the most educational part of the OSTE experience. Here, OSTE leaders act as facilitators of discussions, rather than as experts on the topic. It will serve OSTE leaders/facilitators well to appreciate that all participants, but especially the more experienced faculty, will have much to contribute to the discussion.

    One way to structure the debriefing session is to begin by exploring the group's emotional reaction. How was the experience for the participants? Leaders should allow the group to talk and should keep asking open-ended questions to further the discussion. Next, leaders can go case by case, by first asking one person to summarize and then asking the group to comment on what was challenging or easy. What did the group take away from the experience? At the end of session, leaders should encourage all participants to make a commitment to change one thing the next time they are teaching.

    Once these steps are concluded, you can put it all together to create an OSTE or GOSTE schedule (Fig. 11.1).

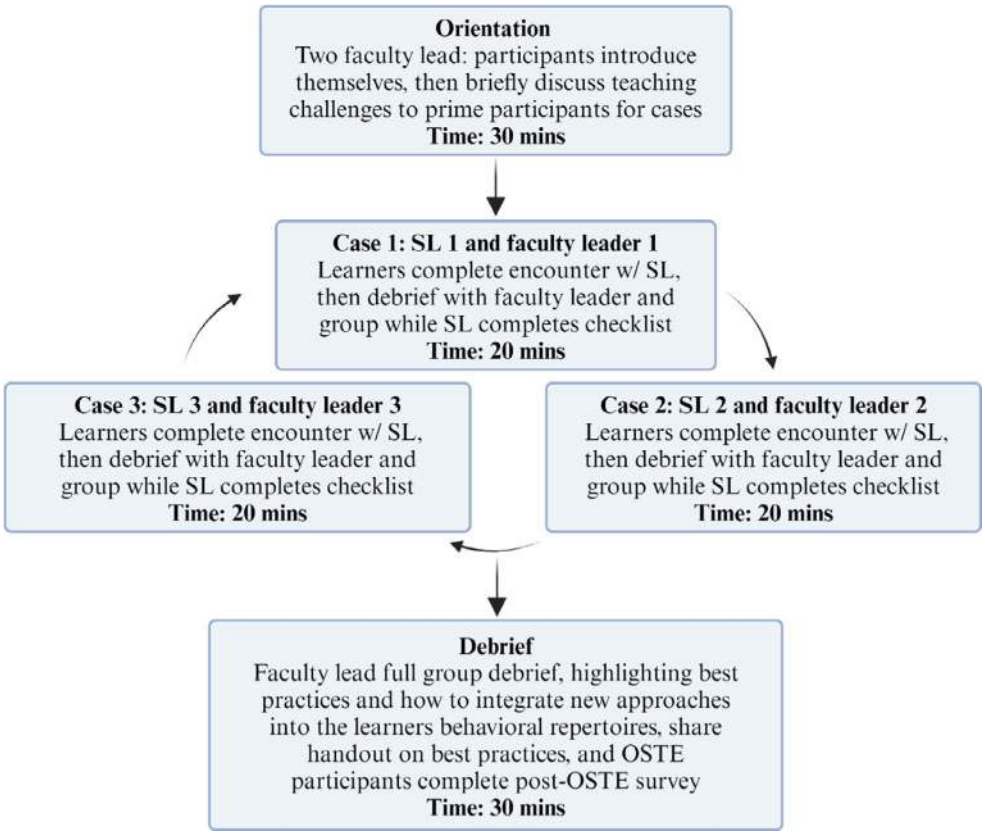


Fig. 11.1 Sample GOSTE schedule (120 min)In this schedule, we would aim to have around nine faculty learners participate in the program. After the orientation, learners would split into groups of three. Group 1 (learners 1, 2, and 3) would begin at Case 1, while groups 2 and 3 begin at the other stations. Learner 1 would take the lead on case 1, while learners 2 and 3 and faculty leader 1 observe and then debrief together. At the next station, a different learner would take the lead. Groups would rotate until they have completed all three cases. Then, all learners would come back together for a full group debrief

An OSTE and GOSTE program at your institution can create a shared mental model for teaching expectations, provide an opportunity to practice and highlight new skills, and build a community of practice for your teaching faculty. The steps in creating and implementing an OSTE are like those for an OSCE with a few unique considerations. There are several reviews [10, 30] on creating and implementing OSTEs which can be helpful to medical educators who want to design impactful OSTEs for their faculty.

Best Practices

1. Create authentic, realistic cases.
2. Clarify the learning skill(s) being assessed.
3. Do a robust needs assessment.
4. Attend to the details: logistics are essential.
5. Prepare (and leave time for) a thoughtful debriefing.
6. Remember that working with faculty is like herding cats! Be prepared for last-minute changes.

References

1. Irby DM, Cooke M, O'Brien BC. Calls for reform of medical education by the Carnegie Foundation for the Advancement of Teaching: 1910 and 2010. Acad Med. 2010;85(2):220–7. https://doi.org/10.1097/ACM.0b013e3181c88449.
2. Leslie K, Baker L, Egan-Lee E, Esdaile M, Reeves S. Advancing faculty development in medical education: a systematic review. Acad Med. 2013;88(7):1038–45. https://doi.org/10.1097/ACM.0b013e318294fd29.
3. Steinert Y, Mann K, Anderson B, et al. A systematic review of faculty development initiatives designed to enhance teaching effectiveness: a 10-year update: BEME Guide No. 40. Med Teach. 2016;38(8):769–86. https://doi.org/10.1080/0142159X.2016.1181851.
4. Alexandraki I, Rosasco RE, Mooradian AD. An evaluation of faculty development programs for clinician-educators: a scoping review. Acad Med. 2021;96(4):599–606. https://doi.org/10.1097/ACM.0000000000003813.
5. Osman C, Bradshaw C, Tewksbury L. Faculty development offered by US Medical Schools: a national survey of pediatric educators. J Contin Educ Heal Prof. 2022;42(3):174–9. https://doi.org/10.1097/CEH.0000000000000403.
6. Liaison Committee on Medical Education. Functions and structure of a Medical School: standards for Accreditation of Medical Education Programs Leading to the MD Degree. Assoc Am Med Coll. Published online 2020.
7. Accreditation Council for Graduate Medical Education. Common Program Requirements (Residency). Accreditation Counc Grad Med Educ. Published online 2021.
8. Accreditation Council for Graduate Medical Education. Common Program Requirements (Fellowship). Accreditation Counc Grad Med Educ. Published online 2021.
9. Fromme HB, Whicker SA, Paik S, et al. Pediatric resident-as-teacher curricula: a national survey of existing programs and future needs. J Grad Med Educ. 2011;3(2):168–75. https://doi.org/10.4300/jgme-d-10-00178.1.
10. Boillat M, Bethune C, Ohle E, Razack S, Steinert Y. Twelve tips for using the objective structured teaching exercise for faculty development. Med Teach. 2012;34(4):269–73. https://doi.org/10.3109/0142159X.2011.599891.

11. Acosta D, Ackerman-Barger K. Breaking the silence: time to talk about race and racism. Acad Med. 2017;92(3):285–8. https://doi.org/10.1097/ACM.0000000000001416.
12. Gonzalez CM, Lypson ML, Sukhera J. Twelve tips for teaching implicit bias recognition and management. Med Teach. 2021;43(12):1368–73. https://doi.org/10.1080/0142159X.2021.1879378.
13. Gonzalez CM, Walker SA, Rodriguez N, Karp E, Marantz PR. It can be done! A skills-based elective in implicit bias recognition and management for preclinical medical students. Acad Med. 2020;95:S150–5. https://doi.org/10.1097/ACM.0000000000003697.
14. Rodriguez N, Kintzer E, List J, et al. Implicit bias recognition and management: tailored instruction for faculty. J Natl Med Assoc. 2021;113(5):566–75. https://doi.org/10.1016/j.jnma.2021.05.003.
15. Acholonu RG, Cook TE, Roswell RO, Greene RE. Interrupting microaggressions in health care settings: a guide for teaching medical students. MedEdPORTAL. 2020;16:10969. https://doi.org/10.15766/mep_2374-8265.10969.
16. Poitevien P, Osman C. Tackling implicit and explicit bias through objective structured teaching exercises for faculty. J Grad Med Educ. 2018;10(3):353–4. https://doi.org/10.4300/JGME-D-17-00906.1.
17. Dembitzer A, Mahowald C, Schaye V, Schwartz C. Tackling challenges in the evaluation of faculty development programs: strategies to assess teaching skills and program impact. Abstracts from the 2022 Annual Meeting of the Society of General Internal Medicine; J Gen Intern Med. 2022;37(S2). https://doi.org/10.1007/s11606-022-07653-8.
18. Trowbridge RL, Snydman LK, Skolfield J, Hafler J, Bing-You RG. A systematic review of the use and effectiveness of the Objective Structured Teaching Encounter. Med Teach. 2011;33(11):893–903. https://doi.org/10.3109/0142159X.2011.577463.
19. Stone S, Mazor K, Devaney-O'Neil S, et al. Development and implementation of an objective structured teaching exercise (OSTE) to evaluate improvement in feedback skills following a faculty development workshop. Teach Learn Med. 2003;15(1):7–13. https://doi.org/10.1207/S15328015TLM1501_03.
20. Ericsson KA, Krampe RT, Tesch-Romer C. The Role of Deliberate Practice in the Acquisition of Expert Performance (Anders Ericsson, Krampe & Tesch-Romer, 1993). Psychol Rev. 1993;100(3):363–406.
21. Ten Cate TJ, Kusurkar RA, Williams GC. How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE Guide No. 59. Med Teach. 2011;33(12):961–73. https://doi.org/10.3109/0142159X.2011.595435.
22. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65(9):S63–7. https://doi.org/10.1097/00001888-199009000-00045.
23. Kirkpatrick D, Kirkpatrick J. Evaluating Training Programs: the Four Levels. 3rd ed. Berrett-Koehler Publishers; 2006.
24. Lu WH, Mylona E, Lane S, Wertheim WA, Baldelli P, Williams PC. Faculty development on professionalism and medical ethics: the design, development and implementation of Objective Structured Teaching Exercises (OSTEs). Med Teach. 2014;36(10):876–82. https://doi.org/10.3109/0142159X.2014.916780.
25. Julian K, Appelle N, O'Sullivan P, Morrison EH, Wamsley M. The impact of an objective structured teaching evaluation on faculty teaching skills. Teach Learn Med. 2012;24(1):3–7. https://doi.org/10.1080/10401334.2012.641476.
26. Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. Acad Med. 1998;73(6):688–95. https://doi.org/10.1097/00001888-199806000-00016.
27. Oh S, Servoss T, Wilkins D. Using the objective structured teaching ecounter to assess resident teaching skills. Fam Med. 2021;53(6):453–6. https://doi.org/10.22454/FamMed.2021.980882.
28. McAndrew M, Eidtson WH, Pierre GC, Gillespie CC. Creating an objective structured teaching examination to evaluate a dental faculty development program. J Dent Educ. 2012;76(4):461–71. https://doi.org/10.1002/j.0022-0337.2012.76.4.tb05278.x.
29. Morrison EH, Rucker L, Boker JR, et al. A pilot randomized, controlled trial of a longitudinal residents-as-teachers curriculum. Acad Med. 2003;78(7):722–9. https://doi.org/10.1097/00001888-200307000-00016.
30. Sturpe DA, Schaivone KA. A primer for objective structured teaching exercises. Am J Pharm Educ. 2014;78(5):1–6. https://doi.org/10.5688/ajpe785104.
31. Srinivasan M, Li STT, Meyers FJ, et al. "Teaching as a competency": competencies for medical educators. Acad Med. 2011;86(10):1211–20. https://doi.org/10.1097/ACM.0b013e31822c5b9a.
32. Osman C, Dembitzer A, Zabar S, Tewksbury L. Using objective structured teaching exercises for faculty development. MedEdPORTAL. 2015:10258. https://doi.org/10.15766/mep_2374-8265.10258.
33. Osman C, Chang L, Tewksbury L. Objective Structured Teaching Exercises (OSTEs) for Faculty: Feasibility and Acceptability. Association of Pediatric Program Directors/Council on Medical Student Education in Pediatrics Joint Meeting; 2013.
34. Howell HB, Desai PV, Altshuler L, et al. Teaching and assessing communication skills in pediatric residents: how do parents think we are doing? Acad Pediatr. 2022;22(2):179–83. https://doi.org/10.1016/j.acap.2021.06.011.
35. Gigante J, Dell M, Sharkey A. Getting beyond "Good job": how to give effective feedback. Pediatrics. 2011;127(2):205–7. https://doi.org/10.1542/peds.2010-3351.

How to Build and Manage an SP Program

12

Brielle Blatt, Virginia Drda, Kathleen Allen, Gene Mamaril, Sean Overstreet, Willie Lombeida, and Katherine Tame

A simulation center's SP program can be a core educational resource. Initially, our simulation activities took place across many sites within our academic center. The building of our simulation center enabled us to centralize simulation education at our institution. This has allowed us to leverage our experience across an array of educational programs and efficiently implement best practices for case development, SP training, and other operational components of simulation. For instance, we provide educational consultation to faculty who are developing new programs, which includes an SP case template (Appendix E) and other support materials. Doing this ensures that the SPs that work in our center become accustomed to the training format and learning process. Using similar cases across programs whenever possible allows our SPs to become highly calibrated to trainee performance and therefore ensures consistency in case portrayal and assessment rating. All programs that run in our center, with few exceptions, use the same behaviorally anchored communication skills as part of the checklist (Appendix F), which further assures highly calibrated SPs and an opportunity for scholarship across programs and developmental levels.

We seek to serve as a resource for our academic community by making education accessible and authentic and ensure our programs meet accreditation expectations. We have regular community meetings with health system and educational stakeholders to debrief past programs, preview and prepare for upcoming programs, share important announcements, and facilitate an open forum to discuss any issues, needs, and goals.

Having a centralized model enables us to maintain integrity of OSCE cases and assessments and execute consistent institutional expectations. Our simulation center staff oversee an OSCE case library and track pertinent details such as programs, case title, authors, learner level, learning objectives, core topics and diagnoses, years completed, organization or department, assessments, and SP demographics. We are able to support new initiatives in health professions education by either building new cases or tailoring existing cases to new expectations. Most recently, we have added OSCE cases and assessments focused on diversity, equity, and inclusion (DEI) learning objectives to our case bank to support simulation-based training in bias mitigation across the educational continuum [1].

This chapter provides the operational framework, based on our experiences, for building and managing a simulation center's SP program.

B. Blatt (✉) · V. Drda · K. Allen · G. Mamaril · S. Overstreet
W. Lombeida · K. Tame
New York Simulation Center for the Health Sciences,
New York, NY, USA
e-mail: brielle.blatt@nyulangone.org

Working with SPs

Building a SP Workforce

When our simulation center first opened, we engaged SPs as independent contractors on a project by project basis using a "studio model" approach. SPs were selected and provided by faculty and coordinators from each department who enrolled the SP as vendors for each short-term SP project. There were benefits and challenges to this approach:

- Advantages included long-standing partnerships and rapport built over time between faculty and coordinators with their selected SPs. SPs were contracted for each assignment with no guarantee of future work, which allowed SP Educators and core faculty to observe SP skills and professionalism on a trial basis. However, there was no cohesive or continuous SP community.
- One of the biggest drawbacks was the issue of SP payment. SP pay rates varied between programs and institutions, and multiple coordinators managed SP invoicing separately, resulting in lengthy delays of SP payments. SP expectations differed between programs, and varying levels of skill and professionalism among SPs posed a challenge to maintaining a high standard of simulation education.

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9_12

Our institution has transitioned to hiring SPs as per diem employees. During this process, we re-defined the SP as "standardized participant" to encompass the wide array of roles simulated in scenarios, including but not limited to patients, nurses, other healthcare professionals, learners, faculty, and family members.

- Advantages include enabling a stronger, more consistent, and unified approach to selecting and hiring SPs with the aim of building long-term business relationships of mutual respect, commitment, and accountability. This investment has been advantageous for our institution and for the SPs, resulting in timely payment and SP retention, as well as fostering community among SPs as a valued part of our team. It also allows us to offer a more equitable distribution of opportunities among our SPs, including a variety of demographic diversity within each program roster. Another benefit was that we were able to advocate for a well-earned hourly wage increase with institutional support of our SPs.

- Challenges during this transition included significant attrition of many long-term SPs who opted out of applying for this new position, a lengthy and detailed onboarding process that new applicants must complete, and enforcing annual employee requirements like vaccination and e-learning compliance module completion by each SP. However, we collaborate closely with Human Resources to assist SPs in completing onboarding tasks and have steadily built a culture of proactive compliance among our SP employees.

Maintaining an SP Database

We maintain a database of our SP employees by gathering self-reported demographic information through a survey of new hires:

- Name and contact information
- Gender identity/pronouns (surveyed annually as this is subject to change over time)
- Age portrayal range (updated every 3 years as age ranges span over a 10-year range)
- Ethnicity/race portrayal
- Languages

We are able to seamlessly select and refer SPs according to demographic needs of cases and curriculum by filtering the SP database. Relevant information regarding SP technical capability and availability for work is also included. The database allows for a unified tracking system of SP professionalism and cancellations, which is monitored and updated as needed.

Hiring: Sourcing, Identifying, Screening, Interviewing, and Onboarding

We follow the Association of Standardized Patient Educators (ASPE) Standards of Best Practices [2] in developing a process to identify SPs and recruit and maintain a cohort that reflects diversity. We collect data to drive our recruitment strategy by identifying demographic gaps in our SP workforce according to gender identity, age, race/ethnicity, and language and by assessing greatest needs based on program curriculum. We conduct outreach to source candidates. We also ask our SPs for recommendations of colleagues, as well as contact local theater companies and network through ASPE. We regularly receive inquiries from individuals interested in SP work and screen hundreds of resumes annually with a keen eye toward past SP, teaching, and related experience. We hire new per diem SPs on a bi-annual basis, prior to peak assessment season.

Our SP Educator identifies which applicants to interview and conducts interviews using objective, specific criteria, including behavioral questions to assess professionalism, skill sets, and trainability. Our academic center's Human Resources specialists manage extending offers and pre-onboarding candidates, per institutional policy. The SP Educator then leads departmental orientations, providing essential information and setting expectations of our SPs. New hires are utilized during their four-month probationary period and receive timely coaching and feedback as needed.

Program-Specific SP Recruitment, Referrals, and Training Processes

When a program seeks to include SPs for simulation, they complete a **Program Request Form** through our website, and an intake meeting is scheduled to discuss program objectives, design and case scenarios. The specific SP demographic needs for each case are communicated to the SP Educator, who recruits and schedules SPs accordingly, securing principal and back-up SPs.

When a faculty or education coordinator requests SP referrals for a specific program that may not be in our simulation center, they complete a **SP Referral Request** through our website, describing the date, location, and demographic needs of each case. The SP Referral Request is forwarded to the SP Educator, who provides the requester with several SP referrals, including SP names and contact information, as well as gender identity and pronouns.

SP training is scheduled by the SP Educator with faculty and SPs to take place 1–2 weeks prior to the first program date. This allows SPs to absorb and integrate the details of the case into their portrayals. Training for case portrayal includes the protocol described in Chap. 2, Step 6, and Table 2.13, with an overview of program logistics and timing, SP expectations, case review and read-through, and rehearsal. SPs practice the case with each other, faculty, and the SP Educator and also practice portraying different learners in the rehearsal encounters.

Faculty may also review and demonstrate physical exam components, when applicable. As new cases are developed, an environment of psychological safety is promoted during training, encouraging and recognizing inter-collaboration of faculty, the SP Educator, and SPs. When applicable, assessment items and ratings are carefully reviewed and practiced with the SPs, as described in Chap. 2, Step 7. When applicable, the format for delivering constructive verbal feedback citing specific describable behavior is reviewed and practiced (Chap. 2, Table 2.16).

SP Quality Assurance

We observe the SPs in person two times during their first year of employment with a prioritization on their probationary period (first four months of employment). We observe one time annually thereafter. We provide specific, constructive, and timely feedback on case portrayal, checklist completion, and verbal feedback on the day of the program. Our SP Educators closely monitor and coach each new SP during their probationary period. We also observe any SPs identified as needing remediation or additional support and spot-check seasoned SPs. Common concerns are accuracy of case portrayal, SPs revealing more information than elicited during the visit, and SP verbal feedback that does not include specific words or behaviors. We also hold annual SP communication checklist and verbal feedback refresher trainings to maintain the quality of assessment and feedback skills.

SP Satisfaction

We strive to retain our valued SP team members by maintaining strong, professional relationships based on mutual respect and communication and by recognizing and acknowledging the nature and demands of SP work. We welcome new SPs to our community at orientation, greet SPs for each program, and offer amenities such as coffee and tea in the SP lounge, as well as changing rooms and lactation spaces. We address SP questions and concerns on an ongoing basis, building a lasting rapport with our SP team.

SIM Center Organizational Structure

Staff Roles and Responsibilities

OSCE teams typically comprise one aspect of an overall simulation center, in addition to simulation (manikin-based SIM), administrative, IT, and operational staff and/or teams. SIM and OSCE teams may have an overlapping role for providing simulation support. Though the granular details of the role's day to day tasks may differ, there is a great deal of overlap regarding the high-level tasks of coordinating program details ahead of time, supporting programs when they occur, and breaking down the program at its conclusion.

Dedicated SIM Center IT staff are a great asset so that there is expertise available to support and troubleshoot issues related to the learning management software, hardware, and AV equipment. Depending on the size of the simulation center, the size of the operations and administrative staff required to function efficiently will vary. Having clearly defined responsibilities among operations and administrative staff is vital regardless of the center's size and programming volume. Please refer to the Appendix T for a sample simulation center organizational chart.

Simulation Operation Associate Onboarding and Training

An effective orientation will set up a Simulation Operation Associate (SOA) or Simulation Technologist for success in the role. It is helpful to immerse them slowly and strategically in various aspects of OSCE planning and operations while providing ample opportunity for them to develop relationships with key OSCE stakeholders. Dedicated training time with IT staff will prove instrumental in establishing a core understanding of the software, hardware, and AV equipment needed to run a successful OSCE program. Please find an example of an SOA orientation schedule in the Appendix U for reference.

Simulation Center Business Operations

Finance: SP Timekeeping/Payroll/Invoice Reconciliation

It is vital to establish a consistent and organized process for tracking the finances associated with a SP program. Once a process is established for tracking SP time worked, it must be communicated with all internal and external stakeholders who schedule SPs for trainings and programs. It is a best practice for SPs to use an organization's timekeeping system

(e.g., Kronos) to ensure their time is captured accurately and that they are paid in a timely manner.

The organization is responsible for establishing a process to invoice internal and external programs for SP wages. Timely invoice reconciliation is integral to maintaining the financial health of an OSCE program, and for this reason, the process should include firm timeframes for sending invoices and payment due dates. The organization will benefit from performing their finance-related tasks and responsibilities consistently and clearly communicating remittance expectations with the stakeholders early on. These business practices will yield accurate and up-to-date bookkeeping while minimizing the amount of time staff spend following up on outstanding invoices. In addition, organizations should consider investing in a payroll system (e.g., QuickBooks) to easily generate and track invoices and payments, which will contribute to the overall quality and efficiency of the invoice reconciliation process.

Simulation Center's OSCE Operations Workflow

Reservation Process

It is essential to have a structured process to receive, evaluate, and support program requests (Table 12.1). Our center's website contains online forms for program reservation requests, change requests, and cancellation requests.

In addition to receiving ad hoc program requests, we also strategically conduct outreach to our stakeholders, as having requests with earlier notice has enabled sound OSCE planning and operations. Approximately 3 months prior to each academic semester, our SP Program Supervisor conducts outreach to new and returning program faculty and coordinators regarding submitting their reservation requests. Since the School of Medicine has a high volume of programs and unique scheduling needs, they submit their reservation requests 1 year in advance.

Day of OSCE Personnel

Although personnel may vary depending on the scale of the program, most programs include a control room operator, hallway monitor, and program faculty or coordinator. Our SOAs may fulfill either the control room operator or hallway monitor role, depending on the needs of the program, and is determined during the intake process. Many programs also include faculty observers to provide immediate verbal feedback and assess learners' skills. We conduct a pre-OSCE huddle to reiterate these established roles, typically 30 minutes prior to the start of the program.

Post Program: Debrief and Incident Reporting

After the program takes place, SOAs send a short "Debrief and Incident" report to SP program directors, Program Manager, and other SOAs. The form includes program name, program date, and report on the following categories: IT, SPs, Equipment and Supplies, Schedule, Feedback from program, and Other. The SOA also logs this report through an online survey tool, enabling us to seamlessly report on issues, identify patterns, and address holistically.

Annual Reports

Annually, we compile a report to calculate the number of programs, sessions (each instance of an education program), number of new and recurrent programs, number of learners, and total hours; this is separated by designated organizations (such as Undergraduate Medical Education, Graduate Medical Education, or Continuing Medical Education) and stakeholders. This enables us to monitor utilization as we develop a strategic plan for expanding and enhancing simulation learning across our health system.

Table 12.1 Example instructions for how to host a program at the SIM Center

Hosting a program at the SIM Center
1. View the SIM Center Calendar to check room availability. An overview of learning spaces is also available on our website.
2. Submit the online Program Request Form. You will receive a confirmation email that your submission was received. • *This form should be submitted no later than 16 weeks before the first date of OSCE programs.*
3. The SIM Center will contact you regarding your submission to confirm dates. The SIM Center will then schedule an intake to discuss logistics and support

OSCE Program Design Timeline: "The Life Cycle of an OSCE"

SP Program Design Timeline

In order to properly plan and execute an OSCE, it is important to have target deadlines for deliverables, as there are many moving parts. We strongly encourage program faculty to submit their reservations at least 16 weeks prior to the first program date, in order to have sufficient operational lead time. Figure 12.1 is a visual of our SIM Center's SP Program Design Timeline.

Intake Process

We conduct intake meetings with program faculty and coordinators to review program needs and details, including space, staff, equipment, logistics, and support. Intakes are facilitated by an SOA or Program Manager, and, depending on the complexity, a SIM Center Director may also attend. See Appendix V for the Intake Program Design Checklist (IPDC), which is used to document information regarding the simulation program for SOAs to follow during set up/planning and to indicate action items for the program and SIM Center team.

Education Consultation

The SIM Center directors provide faculty development for new and existing simulation education programs through formal education consultations. Topics may include scenario design, program structure, task trainer and simulator training, resources and best practices in simulation education, scholarship and SP case development, and program assessment and evaluation options. Faculty who are interested submit a request through our website, and then, the SIM Center Program Associate schedules the meetings and maintains a tracker to collect pertinent details for follow-up and development of program.

Fig. 12.1 SP program design timeline

Technology, Equipment, and Inventory

Desktops, Laptops, and Control Room Equipment

There are two major technological areas for a simulation center. The first is the infrastructure needed to run the learning management system (LMS) and support its features. The second is the equipment for daily operations of OSCEs.

Technological Infrastructure

Table 12.2 details the hardware that makes up the core of the LMS. They are constantly running in the background in order to host the system and enable its features. It allows users to record the encounters during an OSCE, provide assessment capabilities, allow centralized control, and debriefing.

Day of Interfacing Technology

Table 12.3 devices are located in the various areas of the simulation center and are the main interface to the LMS during an OSCE.

Equipment Inventory

There are two categories of equipment needed for OSCES: in-room and on-hand. SOAs manage the equipment inventory, such as replenishing and requesting an order for supplies and requesting additional laundry during periods with a high volume of programs.

In-Room Equipment The equipment in Table 12.4 is included in the standardized setup of each room.

On-Hand Equipment The equipment in Table 12.5 is stored outside of the exam rooms and is available for use based on the set-up needs of each program.

Table 12.2 Core simulation center equipment

Host server		Server that hosts the LMS application
Storage server		Server that hosts all recorded video
Video system	*Pan/Tilt/Zoom (PTZ) Camera(s)*	PTZ camera(s) are located in the Exam Room and are wired back to recording server. These servers temporarily record the audio/video (A/V) feeds and transfer the videos to the storage server.
	Recording server	
Audio system	*Microphone(s)*	The microphone(s) and paging speaker(s) are located in the Exam Room and wired back to the Audio DSP. The DSP filters and adjusts levels for clear sound and is sent to the recording server.
	Paging speaker(s)	
	Audio digital sound processor (DSP)	
A/V control system		A collection of hardware that controls the audio and video routing for the all the Exam Rooms. It primarily consists of touch panels, matrix switchers, and processors.

Table 12.3 Interface technology

Learner interface	A computer located outside of the Exam Room, where the learner logs into the LMS before an encounter. The learner returns to this device to complete their post-encounter note.
Patient interface	A computer located inside the Exam Room, where the SP completes their assessment after an encounter. Sometimes, the SP uses the computer outside the exam room so a faculty observer can enter the room and deliver feedback.
Faculty observation desktop	A computer located in an area outside of the Exam Room opposite of the learner's interface. Faculty members can observe the encounter and complete an assessment.
Control desktop	A computer located in a centralized Control Room. This device connects to the LMS and allows the operator to run the OSCE.
Touch panel	A device located in the Control Room where it allows the operator to select any paging speaker for an announcement.
Paging microphone	A microphone located in the Control Room to allow the operator to announce into any Exam Room.
Multiplexed camera station	A computer connected to a large display located in the Control Room. All the camera feeds in the Exam Rooms will show on this screen. This allows the operator to manipulate camera views for the best shot.

Table 12.4 In-room equipment

☐ Stretcher OR exam table	☐ Otoscope/PanOptic ophthalmoscope
☐ Desktop computer	☐ Gloves (S, M, L)
☐ Phone	☐ Wall clock
☐ Blood pressure cuff	☐ Alcohol prep pads
☐ Sink/soap/sanitizer	☐ Paper towels
☐ Chairs	

Table 12.5 On-hand equipment

☐ Gowns	☐ Headphones
☐ Drapes	☐ Clipboards
☐ Blood pressure cuff (varying sizes)	☐ Stethoscopes
☐ Gloves (additional supply)	☐ Scrap paper
☐ Scrubs (varying sizes)	☐ Tongue depressors
☐ White lab coats (varying sizes)	☐ Applicators
☐ Exam table paper (additional supply)	☐ Reflex hammers/tuning forks
☐ Mayo stands	☐ Exam lights

Conclusion

Efficient, coordinated operations will allow the simulation center and SP program to serve as a core resource for the center's learning community. The use of this framework will help to achieve the vision of your simulation center to serve as an institutional resource for high-quality simulation education and research.

References

1. Blatt B, Allen K, Drda V, Zabar S. Addressing diversity, equity, and inclusion through simulation center education resources. Med Educ. 2022;56(11):1143–4. https://doi.org/10.1111/medu.14900. Epub 2022 Aug 21. PMID: 35989411.
2. Lewis KL, Bohnert CA, Gammon WL, Hölzer H, Lyman L, Smith C, Thompson TM, Wallace A, Gliva-McConvey G. The Association of Standardized Patient Educators (ASPE) Standards of Best Practice (SOBP). Adv Simul (Lond). 2017;2:10. https://doi.org/10.1186/s41077-017-0043-4.

Appendix A. Worksheet: OSCE Planning

OSCE Project Name	
OSCE Goals	
Number and Type of Trainees	
Number and Type of Stations	
Number of SPs needed	
Number of faculty needed	
Potential Timing	
Potential Space	
Approximate Budget	
Potential Funding Sources	
Instructional Strategies	

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Appendix B. Worksheet: OSCE Budget

Budget items to consider	Amount needed	In kind	Cost/learner
SPs—training and OSCE performance Check for local rates; costs vary depending on location and simulation task. Factor early arrival and debriefing/ de-roling time into payment.	$___ /hour x ____ SPs x ____ hours (training + OSCE sessions) **=** _____		
Faculty observers (feedback and prebrief/debrief)			
Medical supplies Does not need to be sterile, but should be authentic			
Office supplies Printing paper evaluation forms, pens			
Refreshments for learners, SPs, and faculty			
Recording equipment			
Data entry and report assembly (individual learner data and aggregate program data) May be performed Learning Management System or staff			
Data analysis Faculty/staff with statistical analysis capabilities are vital for scholarship and longitudinal tracking			
Other (i.e. Space)			
Total			

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Appendix C. Worksheet: Breakdown of OSCE Responsibilities

OSCE project name:		Date:	
Tasks		Individuals involved	Deadlines
Initial planning			
3–4 months before the OSCE	Decide on format (e.g., time frame, modality, number of stations)		
	Create a blueprint (identify competencies to be assessed)		
	Identify appropriate OSCE location (stations and assembly rooms) or online platform		
	Recruit staff (for administrative tasks, monitoring, time keeping)		
	Communicate with learners (provide dates/times, explain format and procedure)		
	Clarify and negotiate budget (e.g., SP costs, refreshments)		
	Consider videotaping and arrange for set-up		
	Decide on what stations to maintain from previous OSCEs/develop new/import from station banks		
	Decide on SP/rate recruitment and schedule		
Station/material preparations			
1 week–3 months before the OSCE	Determine SP payment process		
	Make room arrangements and/or determine online platforms		
	Recruit faculty for pre-brief, debrief and observation		
	Prepare station materials (i.e., develop new ones, adapt old ones if appropriate)		
	Recruit SPs		
	Prepare faculty (e.g., circulate station/format information, feedback training)		
	Prepare props (e.g., fake pill bottles, charts)		
	Train SPs		
General on-site and/or online Preparations			
1–2 weeks before the OSCE	Order supplies (e.g., paper, folders)		
	Assign SPs, faculty and learners (create assignment sheets)		
	Develop rotation schedules (i.e., matrix of learners, rotations and stations; if online it will be useful to include pictures of learners to double check that the right person is in the right station)		
	Prepare invoices and necessary paperwork for SP payment		
	Print station materials		
	Prepare name tags/labels for learners (assign numbers to learners)		
	Prepare signs (e.g., station numbers, arrows to signal flow)		
	Orient hall monitors and timekeepers		
	Practice walk-through		

(continued)

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

OSCE project name:		Date:	
Tasks		Individuals involved	Deadlines
OSCE administration			
Day of the OSCE	Prepare stations and assembly rooms (e.g., signs, station materials, refreshments) or set up breakout rooms on online platform (e.g., stations, SP, faculty and learner assembly rooms, orientation and debriefing rooms)		
	Assign substitutes (if necessary)		
	Orient faculty, SPs, and other personnel (in-person and/or online)		
	Position faculty, SP, hall monitors, timekeepers (in-person and/or online)		
	Orient learners (in-person and/or online)		
	Guide learners to individual starting stations		
	Time stations and make announcements (start, feedback, station changes, end of OSCE)		
	Manage emergencies (e.g., equipment/Wi-Fi breakdown)		
	Assure smooth changeovers of SPs, faculty, learners (in-person and/or online)		
	Assemble learners for the OSCE group debriefing (include faculty if possible)		
	Assemble SPs for debriefing and de-rolling		
	Collect and count all forms (check completions if submitted online)		
	Clean up stations and assembly rooms		
Post-OSCE tasks			
Days to weeks after OSCE	Debrief with organizers (if possible, include timekeepers, SP trainers and others involved in the administration)		
	Sort out forms (in-person and/or online)		
	Ensure timely SP payment		
	Enter data and evaluation results if not completed through an electronic Learning Management System		
	Analyze data		
	Report evaluation data (e.g., report cards)		
	Report on experience internally and externally (e.g., presentations, articles)		

Appendix D. Worksheet: OSCE Blueprint

Station			Competencies					Implementation Notes	SP Characteristics
Case number and name	Case objectives	Content Areas	Core Comm Skills	Hx Gather	Educ & Coun	Tx plan & Man	Case-Specific Skills		

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Appendix E. Template Materials for OSCE Case/Station Development

Template for developing materials for an OSCE station: station overview, participant instructions, and SP instructions.

Station Overview: Case/Station Name		
Development date		
Station developer(s) and contact information		
Learners (intended and potential)		
Objectives	To test learners' ability to: 1. 2. 3. 4.	
Logistics	Personnel:	
	Forms:	• Resident instructions • SP instructions • SP rating form • Faculty rating form
	Room requirements/ resources:	• • •

Participant Instructions: Case/Station Name	
Patient information	Name: Age: Occupation:
Reason for encounter	
Background	• • • • •
Your role	
Your tasks	• • • (Indicate if a physical exam is expected)

SP Instructions: Case/Station Name		
Scenario	Your name is … (How did the current encounter come about?)	
History of present illness	Chief complaint:	(Reason for visit)
	Where	(Location and radiation of symptom)
	When	(When it began, fluctuation over time, duration)
	Quality	(What it feels like)
	Quantity	(Intensity, extent, degree of disability)
	Aggravating/alleviating factors	(What makes it better/worse)
	Associated symptoms	(Other manifestations)
	Beliefs	(What does the patient think is wrong)

(continued)

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

SP Instructions: Case/Station Name	
Current life situation	(Where does patient live/work, …)
Personality	(Key emotional tone and approach to responses)
Past medical history	(Past illnesses including surgical or psychiatric conditions)
Family medical history	(Past medical, surgical, and/or psychiatric conditions relevant for the case)
Medications	(List with quantity if relevant)
Allergies	(List)
Social history	(e.g., smoking, drugs, alcohol, diet, exercise) _____
Encounter beginning	(What SP should do at the beginning of the encounter, opening statement)
Middle	(How the SP should respond with information and emotionally given various learner approaches)
End (2-minute warning)	(How the SP should allow the learner to close the encounter)
Physical exam	(How the SP should react to relevant physical exam maneuvers, what the participant will be looking for)

Appendix F. Worksheet: SP Rating Form Development (OSCE)

The following checklist template can be adapted for any case by adding case-specific items and anchors and removing any non-applicable sections (e.g., physical exam).

[place participant ID here]

Case name: _____

SP Name:

Communication

Information gathering	**Not done**	**Partially done**	**Well done**	*Comments*
Elicited your responses using appropriate questions: • No leading questions • Only 1 question at a time	Impeded story by asking leading/judgmental questions AND more than one question at a time	Used leading/judgmental questions OR asked more than one question at a time	Asked questions one at a time without leading patient in their responses	
Clarified information by repeating to make sure he/she understood you on an ongoing basis	Did not clarify (did not repeat back to you the information you provided)	Repeated information you provided but did not give you a chance to indicate if accurate	Repeated information and directly invited you to indicate whether accurate	
Allowed you to talk **without interrupting**	Interrupted	Did not interrupt directly BUT cut responses short by not giving enough time	Did not interrupt AND allowed time to express thoughts fully	
Relationship development	**Not done**	**Partially done**	**Well done**	*Comments*
Communicated concern or intention to help	Did not communicate intention to help/concern via words or actions	Words OR actions conveyed intention to help/concern	Actions AND words conveyed intention to help/concern	
Nonverbal behavior enriched communication (e.g., eye contact, posture)	Nonverbal behavior was negative OR interfered with communication	Nonverbal behavior demonstrated attentiveness	Nonverbal behavior facilitated effective communication	
Acknowledged emotions/ feelings appropriately	DID NOT acknowledge emotions/feelings	Acknowledged emotions/ feelings	Acknowledged and responded to emotions/ feelings in ways that made you feel better	
Was **accepting/ nonjudgmental**	Made judgmental comments OR facial expressions	Did not express judgment but did not demonstrate respect	Made comments and expressions that demonstrated respect	
Used words you understood and/or explained **jargon**	Consistently used jargon WITHOUT further explanation	Sometimes used jargon AND did not explain it	Explained jargon when used OR avoided jargon completely	

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Patient education and counseling	**Not done**	**Partially done**	**Well done**	*Comments*
Asked questions to see what you understood (check your understanding)	Did not check for understanding	Asked if patient had any questions BUT did not check for understanding	Assessed understanding by checking in throughout the encounter	
Provided **clear explanations**/ information	Gave confusing OR no explanations which made it impossible to understand information	Information was somewhat clear BUT still led to some difficulty in understanding	Provided small bits of information at a time AND summarized to ensure understanding	
Collaborated with you in identifying possible next steps/ plan	Told patient next steps/plan	Told patient next steps THEN asked patient's views	Told patient options THEN mutually developed a plan of action	

Case-Specific Skills

Competency: _____	**Not done**	**Partially done**	**Well done**	*Comments*
1.				
2.				
3.				
Competency: _____	**Not done**	**Partially done**	**Well done**	*Comments*
1.				
2.				
3.				
Competency: _____	**Not done**	**Partially done**	**Well done**	*Comments*
1.				
2.				
3.				

Physical Exam

Competency: _____	**Not done**	**Partially done**	**Well done**	*Comments*
1.				
2.				
Competency: _____	**Not done**	**Partially done**	**Well done**	*Comments*
1.				
2.				

Patient Satisfaction

The doctor (resident)….	**Not done**	**Partially done**	**Well done**	*Comments*
Fully explored my **experience of the problem** (concerns, symptoms, functions, feelings, ideas)	Did not explore	Explored some aspects of my experience but not all	Fully explored major aspects of my experience	
Explored my **expectations about visit** (problem, solution)	Did not explore	Partially explored my expectations	Fully explored my expectations	
Took a **personal interest** in me; treated me as a **person**	Did not see me as a person	Viewed me as a person, but did not take personal interest	Took an active personal interest in me	
Gave me **enough information**	I was not given anywhere close to enough information	I was given some information, but I still had questions	I was given all the information I wanted/ needed	

Patient Activation

This encounter….	Not done	Partially done	Well done	*Comments*
This encounter helped me to **understand the nature and causes** of my problem	Did not help me understand	Helped me understand some things but not everything	Helped me fully understand what happened	
After the encounter, I **knew and understood the different medical treatment options available**	I did not find out about treatment options	I found out about some of the treatment options	I found out about all of the relevant treatment options	
This visit made me feel **confident I can keep my problem interfering** too much with my life	Did not affect my confidence	Helped me feel more confident that I could keep my health problem from interfering w/ life	Helped me feel very confident that I could keep my health problem from interfering w/ life	
Because of this encounter, **I am confident I can figure out solutions if something new** comes up	Did not affect my confidence	Helped me feel somewhat confident that I could deal with new issues	Helped me feel quite confident that I could deal with new issues	

Would you recommend this doctor to a friend or family member for his/her….?

Communication skills:	**Not recommend**	**Recommend with reservations**	**Recommend**	**Highly recommend**

Overall, how would you rate this doctor's professionalism?

	Not at all professional	**Somewhat professional**	**Professional**	**Very professional**
Professionalism	***Most*** *of the following* • Disrespectful • Not compassionate • Not accountable • Not sensitive/responsive to my needs/situation	*A* ***few*** *of the following* • Disrespectful • Not compassionate • Not accountable • Not sensitive/responsive to my needs/situation	***3*** *of the following* • Respectful • Compassionate • Accountable • Sensitive/responsive to my needs/situation	***All*** *of the following* • Respectful • Compassionate • Accountable • Sensitive/responsive to my needs/situation

SPECIFIC (< 1 min) FEEDBACK:
COMMENTS (additional remarks, factors affecting your score, impressions)

Appendix G. Sample OSCE Case Materials: Medical Error Disclosure

Station overview, participant (resident) instructions, and SP instructions for the "Medical Error Disclosure" case are included below.

Station Overview: Medical Error Disclosure		
Development date	April 1, 2012 (updated, 2022)	
Station developers	S. Zabar, E. Kachur, K. Hanley, A. Kalet	
Learners (intended and potential)	General Internal Medicine residents Also suitable for: Medicine residents, medical students	
Objectives	To test learners' ability to: 1. Admit an error has been made 2. Express empathy 3. Address patient concerns surrounding an error 4. Reestablish rapport	
Logistics	Personnel:	Standardized patient, male, early forties, dressed in street clothing (casual), sitting in chair.
	Room requirements/ resources:	• 2 chairs • Medical equipment—None • Other props—None

Resident Instructions: Medical Error Disclosure	
Patient information	Name: John McCoy Age: 42 Marital status: Single, never married Occupation: Musician/waiter
Reason for encounter	Patient initiated visit for recent onset of tiredness and frequent urination
Background	• 4 months ago, you saw this patient in clinic for a check-up. He had no complaints. You performed a complete history and physical. • His FMH was significant for diabetes and high cholesterol. • At the last visit, you ordered electrolytes and cholesterol and told him to make a follow-up appointment to review his labs 2 weeks later. He cancelled the appointment. • As you review his EMR before the visit, you notice his sugar was 190 on the lab work you ordered 4 months ago. There is nothing in the record to suggest that you responded to this abnormality. • Today, he presents complaining of fatigue, excessive urination, and thirst. • His finger stick is 250 and urine dip has glucose.
Your role	Resident on ambulatory care rotation
Your tasks	• Explain the follow-up error • Explore the patient's current problem • Develop a plan • DO NOT CONDUCT A PHYSICAL EXAM

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

SP Instructions: Medical Error Disclosure		
Scenario	Your name is John McCoy. You are 42 years old and single. You are a jazz pianist but have been waiting tables in order to make ends meet.	
	Four months ago, you came to the clinic for the first time. You realized that you had not been to a doctor for a while and should have a check-up. Your mom has diabetes and takes tons of medications, and your father has high cholesterol and is crazy about what he eats. While you keep a pretty healthy lifestyle in terms of diet and exercise and consider yourself too young to have problems, you thought your family history might put you at risk so you decided to get checked out.	
	You liked the doctor you saw the last time you were here. You felt they listened and took your concerns seriously. They ordered tests to check your sugar and cholesterol. You got your blood drawn fasting (as the doctor requested) and made a follow-up appointment. Unfortunately, you had to cancel it at the last minute because of an unexpected music gig (these don't come across as often as you'd like). When you called to cancel, the computer system was down, and the receptionist said that you would have to call back. When you called again later, you got a busy signal and gave up trying to get through. No one from the clinic called you back to reschedule, so you figured that the results must have been fine anyway.	
	About 2 weeks ago you started to feel really exhausted. You noticed that you were constantly thirsty and going to the bathroom a lot. You started to keep a plastic liter bottle of water with you at all times and needing to refill it frequently. You are urinating constantly, at high volume but without any associated pain. It seems like all of your clothes are little loose and you wonder whether you've also lost a few pounds (you are of medium build; your weight is usually steady).	
	You have now missed several days of work because the constant bathroom runs were really disruptive to your work in the restaurant. You had someone cover for you at your job, but you were afraid that you would not get your day shift back. Yesterday, you decided to work anyway even though you seem to be running to the bathroom every 10 min. By the end of the day, you were totally exhausted and anxious.	
	For the last 4 days you have been calling the clinic daily trying to make an appointment with the same physician you saw the last time. Finally, today, the doctor is able to see you. You are feeling exhausted, a little worried about what is wrong with you and frustrated with the clinic appointment system. Today you've already had to wait more than an hour before getting to see the doctor. You are expected at your job shortly.	
History of present illness	Chief complaint:	Tiredness, frequent urination
	Where	General exhaustion
	When	Problem started about 2 weeks ago, you are going to the bathroom every 10 minutes, waking frequently at night
	Quality	Exhausted due to lack of sleep. No pain on urination.
	Quantity	High volume urination
	Aggravating/alleviating factors	None
	Associated symptoms	Increased thirst, possible weight loss, no fever
	Beliefs	Fear it may be diabetes like your mother has
Current life situation	You moved to New York City when you were 20 and have been living with a roommate in the East Village for the last 10 years. You have never been married and have no children, but you have had a number of steady girlfriends and are usually in a relationship. Your parents live in Ohio, as does your 2-year younger brother, who is married.	
	You work as a waiter at the Café Wha and intermittently play jazz piano with various local groups. You are still hoping to make it as a pianist, but it hasn't worked out that well so far. Your financial situation is slightly unstable which can put you on edge at times.	
	Over the last few months, nothing much has changed in your life. You job has been as stressful as always. No changes in your diet or exercise. You have recently started to become worried that you might have diabetes.	
Personality	You tend to be a little dramatic. When you are happy, you are very upbeat, and when you are upset, you can get quite angry, raising your voice (although not shouting or swearing).	
Past medical history	Common childhood diseases and colds—otherwise unremarkable. No past surgical or psychiatric history.	
Family medical history	Your mother has had diabetes for about 15 years (you think she was diagnosed at age 45). You know that it has been getting worse and fear that she does not take great care of herself.	
	You father has high cholesterol and obsessed with his health. He is always on a special diet and cooks all his own food. In your opinion, your parents have a strange relationship, but it seems to work for them.	
Medications	None	
Allergies	None	
Social history	You do not smoke and do not use recreational drugs. You drink alcohol at least three times per week, usually having two to three drinks each time.	
	You are sexually active (using condoms for protection) with a girlfriend whom you have had for the past 6 months.	
	Normally, you eat and sleeping well (when work allows) and stay active by going to the gym occasionally. Of course, your restaurant job also keeps you moving constantly.	

(continued)

SP Instructions: Medical Error Disclosure	
Encounter beginning	**When the resident enters the room**, you are sitting in a chair in the exam room talking to a waiter colleague on a cell phone, trying to get someone to cover for you as you may be late for work today since you are still at the doctor's office. You are upset, interrupting the person on the other end of the phone line, and end the conversation about 20 seconds after the resident enters the room. When you hang up, you are still upset about having to miss work although you are glad to finally see the physician again. You express some frustration about not getting an appointment sooner, waiting so long in the waiting room which results in more work problems and never hearing back about the test results.
	If asked in an open-ended way why you are here, state: "I'm feeling really awful, I am peeing all the time, and I never found out what my blood tests showed."
	Provide details about your current state as indicated above.
	How have you been before problem started?—"Fine, I guess. I have been busy at work and trying to get as many music gigs as I can. Maybe I was a little thirsty. I think I lost a few pounds, but who pays attention. I don't really keep regular hours."
Middle	**If/when you are told a mistake was made** (i.e., the fasting blood tests you did 4 months ago showed you had diabetes, all other labs were normal), regardless of where it occurs in the interview, take a moment to let it set in and then respond with anger. Raise your voice (but don't shout), look the resident straight in the eye and state: "So I had diabetes 4 months ago? Why didn't anyone call me? Is this what is going on now?" and "Could we have avoided all this? What's going on here? I had to miss days of work because of this! Will there be any long-term damage because you did not catch this earlier? You mean I could have been dying, and no one would have told me my blood sugar was high unless I came here. Aren't there systems to contact patients with abnormal tests? I assumed no news was good news."
	When you realize the **long-term damage will be nil or minimal**, you become a little less agitated but state in a frustrated way: "Why did this happen? What if this was something really serious? I mean, my God, does this happen here all the time? There is something wrong with the system here!"
	If the resident remains **apologetic and nonconfrontational**, you calm down a little and ask: "Well, when can I go back to work? How am I going to get better?"
	If the resident acknowledges that a mistake was made, but then becomes **defensive**, **does not empathize**, or say they are sorry or makes up a bizarre story, get more upset: "I mean, me missing work today would have been totally unnecessary right? If you guys actually did your job, I wouldn't have had to get so sick. I knew I shouldn't have come to this clinic." If the resident blames you for not rescheduling explain that when you called to cancel the computer system was down and no one contacted you when it was up again. Thus, you surmised that it was not so important.
	Ongoing: If the Resident is empathic/truthful/straightforward, become more and more calm. If the Resident is defensive or evasive, become more and more upset/angry.
End (2-minute warning)	**Regardless of resident's reactions**, calm down a bit and stop additional questions about how this happened to allow the interviewer close down the encounter. State: "Well, so I have diabetes. Now what?"
	If the resident is acting appropriately, you calm down in response to the effective interventions. If the resident acted inappropriately, mention that you intend to take up the problem with someone else. Cross your arms and avoid eye contact but stop making angry comments.

Appendix H. Sample SP Rating Form (OSCE)

The following is an SP checklist for the "Medical Error Disclosure" case found above in Appendix G.

Resident ID #

Medical Error Disclosure

SP Name:

Communication

Information gathering	**Not done**	**Partially done**	**Well done**	*Comments*
Elicited your responses using appropriate questions: • No leading questions • Only 1 question at a time	Impeded story by asking leading/judgmental questions AND more than one question at a time	Used leading/judgmental questions OR asked more than one question at a time	Asked questions one at a time without leading patient in their responses	
Clarified information by repeating to make sure he/she understood you on an ongoing basis	Did not clarify (did not repeat back to you the information you provided)	Repeated information you provided but did not give you a chance to indicate if accurate	Repeated information and directly invited you to indicate whether accurate	
Allowed you to talk **without interrupting**	Interrupted	Did not interrupt directly BUT cut responses short by not giving enough time	Did not interrupt AND allowed time to express thoughts fully	
Relationship development	**Not done**	**Partially done**	**Well done**	*Comments*
Communicated concern or intention to help	Did not communicate intention to help/concern via words or actions	Words OR actions conveyed intention to help/concern	Actions AND words conveyed intention to help/concern	
Nonverbal behavior enriched communication (e.g., eye contact, posture)	Nonverbal behavior was negative OR interfered with communication	Nonverbal behavior demonstrated attentiveness	Nonverbal behavior facilitated effective communication	
Acknowledged emotions/feelings appropriately	DID NOT acknowledge emotions/feelings	Acknowledged emotions/feelings	Acknowledged and responded to emotions/feelings in ways that made you feel better	
Was **accepting/nonjudgmental**	Made judgmental comments OR facial expressions	Did not express judgment but did not demonstrate respect	Made comments and expressions that demonstrated respect	
Used words you understood and/or explained **jargon**	Consistently used jargon WITHOUT further explanation	Sometimes used jargon AND did not explain it	Explained jargon when used OR avoided jargon completely	
Patient education and counseling	**Not done**	**Partially done**	**Well done**	*Comments*
Asked questions to see what you understood (check your understanding)	Did not check for understanding	Asked if patient had any questions BUT did not check for understanding	Assessed understanding by checking in throughout the encounter	

(continued)

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Provided **clear explanations/** information	Gave confusing OR no explanations which made it impossible to understand information	Information was somewhat clear BUT still led to some difficulty in understanding	Provided small bits of information at a time AND summarized to ensure understanding	
Collaborated with you in identifying possible next steps/plan	Told patient next steps/ plan	Told patient next steps THEN asked patient's views	Told patient options THEN mutually developed a plan of action	

Case-Specific Skills

Delivering bad news	**Not done**	**Partially done**	**Well done**	*Comments*
Prepared you to receive the news: • Entered room prepared to deliver news • Ensured sufficient time and privacy	Entered room in a manner unfitting the news AND physically situated self far away	Entered room in a manner unfitting the news OR physically situated him/herself far from you	Entered room in a manner befitting the news AND physically situated him/ herself close to you	
Gave a **warning shot** (e.g., "I have some bad news for you…")	No warning shot	Attempted to deliver warning shot, BUT did so inappropriately (did not pause for your assent OR warning shot too long)	Gave you a well-timed warning shot	
Gave you **opportunity to respond:** • Remained sensitive to your venting of shock, anger, disbelief, accusations • Attended to your emotions before moving on	Responded inappropriately to your emotional reaction (no opportunity to vent, cut you off, became defensive)	Allowed you to emotionally respond (vent) BUT did not address/acknowledge response before moving on	Allowed you to express your feelings, fully giving you the feeling you were being listened to before moving on	
Directly asked what you are feeling: "What are you thinking/feeling?"	Did not ask specifically "What are you thinking/ feeling?"	Acknowledged your feelings (e.g., "I see that you are upset…") BUT did not ask you to name your emotions	Specifically asked you "What are you thinking/ feeling?"	
Managing a difficult situation	**Not done**	**Partially done**	**Well done**	*Comments*
Avoided **assigning blame**	Became defensive/ argumentative AND assigned blame to a person/ department	Became defensive/ argumentative OR assigned blame to a person/department	Remained calm AND did not mention blame someone else	
Maintained professionalism by **controlling emotions**	Unable to control emotions, became dismissive, defensive or/and condescending	Attempted to control emotions (e.g., only somewhat dismissive, defensive or condescending)	Maintained high level of professionalism, no defensiveness, anger, frustration	
Disclosure and accountability	**Not done**	**Partially done**	**Well done**	*Comments*
Disclosed error • Direct (used the words "error" or "mistake") • Prompt disclosure	Did not directly disclose the error (there was a "problem") NOR was the explanation upfront	Did not directly disclose the error (there was a "problem") OR directly disclosed late in the interview	Directly disclosed the error upfront	
Personally **apologized** for the error ("I am sorry that this happened")	Did not apologize for error NOR for the inconvenience it caused you	Apologized for the error OR for the inconvenience it caused you	Apologized for the error AND for the inconvenience it caused you	
Shared the **cause of the error** (i.e., explained issues with system)	Did not acknowledge issues with system	Acknowledged issue with system BUT was dismissive/ condescending	Acknowledged issue with system AND was genuine in addressing it	

(continued)

Took **responsibility** for situation	Took no personal responsibility for your present situation (e.g., assigns your problem to other person/department)	Took a general responsibility as part of the department for your present situation	Took a personal responsibility for your situation ("I will…)	
Identified future **preventive strategies** to prevent situation from happening again	Did not address how situation would be prevented in future	Made general suggestion for improvement (e.g., "We'll look into it," "I'll make a note of it to my Attending")	Offered specific strategies for potential improvement of system	

Patient Satisfaction

The doctor (resident)….	**Not done**	**Partially done**	**Well done**	*Comments*
Fully explored my **experience of the problem** (concerns, symptoms, functions, feelings, ideas)	Did not explore	Explored some aspects of my experience but not all	Fully explored major aspects of my experience	
Explored my **expectations about visit** (problem, solution)	Did not explore	Partially explored my expectations	Fully explored my expectations	
Took a **personal interest** in me; treated me as a **person**	Did not see me as a person	Viewed me as a person, but did not take personal interest	Took an active personal interest in me	
Gave me **enough information**	I was not given anywhere close to enough information	I was given some information but I still had questions	I was given all the information I wanted/needed	

Patient Activation

This encounter….	**Not done**	**Partially done**	**Well done**	*Comments*
This encounter helped me to **understand the nature and causes** of my problem	Did not help me understand	Helped me understand some things but not everything	Helped me fully understand what happened	
After the encounter, I **knew and understood the different medical treatment options available**	I did not find out about treatment options	I found out about some of the treatment options	I found out about all of the relevant treatment options	
This visit made me feel **confident I can keep my problem interfering** too much with my life	Did not affect my confidence	Helped me feel more confident that I could keep my health problem from interfering w/ life	Helped me feel very confident that I could keep my health problem from interfering w/ life	
Because of this encounter, **I am confident I can figure out solutions if something new** comes up	Did not affect my confidence	Helped me feel somewhat confident that I could deal with new issues	Helped me feel quite confident that I could deal with new issues	

Would you recommend this doctor to a friend or family member for his/her….?

Communication skills:	**Not recommend**	**Recommend with reservations**	**Recommend**	**Highly recommend**
Medical competence:	**Not recommend**	**Recommend with reservations**	**Recommend**	**Highly recommend**

Overall, how would you rate this doctor's professionalism?

	Not at All professional	Somewhat professional	Professional	Very professional
Professionalism	***Most*** *of the following* • Disrespectful • Not compassionate • Not accountable • Not sensitive/ responsive to my needs/situation	*A* ***few*** *of the following* • Disrespectful • Not compassionate • Not accountable • Not sensitive/ responsive to my needs/situation	***3*** *of the following* • Respectful • Compassionate • Accountable • Sensitive/responsive to my needs/situation	***All*** *of the following* • Respectful • Compassionate • Accountable • Sensitive/responsive to my needs/situation

SPECIFIC (< 1 min) FEEDBACK:
COMMENTS (additional remarks, factors affecting your score, impressions)

Appendix I. Sample Faculty Rating Form

The following form (details tailored to the "Medical Error Disclosure" case, Appendix G) is intended to assist faculty observers in evaluating OSCE participant (resident, in this case) performance.

Communication

Information gathering	**Does not meet expectations**	**Meets expectations**	**Exceeds expectations**	*Comments*
	1. Elicited patient responses using appropriate questions (no leading questions, only 1 question at a time). 2. Clarified information by repeating to make sure he/she understood patient on an ongoing basis. 3. Allowed patient to talk without interrupting.			

Relationship development	**Does not meet expectations**	**Meets expectations**	**Exceeds expectations**	*Comments*
	1. Communicated concern or intention to help. 2. Nonverbal behavior enriched communication (eye contact, posture). 3. Acknowledged emotions/feelings appropriately. 4. Was accepting/nonjudgmental. 5. Used words patient understood and/or explained jargon.			

Patient education and counseling	**Does not meet expectations**	**Meets expectations**	**Exceeds expectations**	*Comments*
	1. Asked questions to see what patient understood (checked for understanding). 2. Provided clear explanations/information. 3. Collaborated with patient in identifying possible next steps/plan.			

Case-Specific Skills

Disclosing error	**Does not meet expectations**	**Meets expectations**	**Exceeds expectations**	*Comments*
	1. Used "breaking bad news format", i.e., prepared patient, gave warning shot, was unambiguous in delivery, gave patient an opportunity to respond. 2. Was prompt and direct with disclosure ("I made a mistake"), personally apologized, and took responsibility for next steps. 3. Shared cause of error, and let patient know what would be done to prevent error from happening again.			

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Managing a difficult situation	**Does not meet expectations**	**Meets expectations**	**Exceeds expectations**	*Comments*
	1. Maintained professional composure and controlled emotions. 2. Avoided assigning blame—to someone else within "the system" or to the patient.			

Based on the resident's performance in this case, how much supervision would the resident need to handle this case in actual practice (based on Entrustable Professional Activity rating):

*Faculty "entrustment" rating**	**Requires direct supervision (i.e., supervisor present)** **1**	**Requires indirect supervision (i.e., preparing ahead with supervisor)** **2**	**Ready for unsupervised practice** **3**	**Can supervise others** **4**
Comments				

**This item will be the only item collected for assessment*

Additional Comments:

Appendix J. Worksheet: OSCE Participant Rotation Schedule

This worksheet includes rotation timeslots (columns) for ten participants in an eight-station OSCE including two rest stations. To complete the schedule, first fill in the names of the participants in the left-hand column. Add the start times of each rotation across the top blank row (i.e., if the OSCE begins at 1:00 pm and each learner spends 18 minutes at each station, you would write 1:00 pm, 1:18 pm, 1:36 pm, etc. For the first learner, write the station numbers in order (i.e., 1–4, rest, 5–8, rest). Do the same with the next row, except shift the order of the stations by one (i.e., begin with rest and end with station #8), and continue completing the table in this manner until all the rows are filled. **Refer to Table 2.18 for a sample completed schedule.**

	Date: Brief orientation from: _____. Each learner has ___ encounters and ___ rest stops. Last encounter ends at _____. Faculty and resident debrief from: _____.									
	Timing									
Learner name										
	1	2	3	4	Rest	5	6	7	8	Rest
	Rest	1	2	3	4	Rest	5	6	7	8
	8	Rest	1	2	3	4	Rest	5	6	7
	7	8	Rest	1	2	3	4	Rest	5	6
	6	7	8	Rest	1	2	3	4	Rest	5
	5	6	7	8	Rest	1	2	3	4	Rest
	Rest	5	6	7	8	Rest	1	2	3	4
	4	Rest	5	6	7	8	Rest	1	2	3
	3	4	Rest	5	6	7	8	Rest	1	2
	2	3	4	Rest	5	6	7	8	Rest	1

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Appendix K. Survey: Participant Evaluation of OSCE Experience (designed for a residency OSCE; PGY = post-graduate year)

Please indicate your response to each of the questions listed in the columns:

		How much *exposure* have you had to similar cases?			How *difficult* was this case for you?			How much did you *learn* from doing this case?			How would you rate your *overall performance* in this case?				
Station #	**Case Name**	**None**	**Some**	**A lot**	**Too easy**	**Just right**	**Too hard**	**Nothing**	**Some**	**A lot**	**Poor**	**Fair**	**Good**	**Excel-lent**	**What did you think was *the point* of this case?**
1		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
2		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
3		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
4		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
5		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
6		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
7		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
8		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

Please indicate how much you agree or disagree with each of the items below:

In general, this OSCE…		**Strongly disagree**	**Somewhat disagree**	**Somewhat agree**	**Strongly agree**
1	Helped me identify my strengths and areas I need to improve	☐	☐	☐	☐
2	Stimulated me so that I'll go and learn more about some of the topics covered	☐	☐	☐	☐
3	Taught me something new	☐	☐	☐	☐
4	Provided me with valuable feedback	☐	☐	☐	☐
5	Was a lot like real-life clinical encounters	☐	☐	☐	☐
6	Evaluated my skills fairly	☐	☐	☐	☐
7	Was enjoyable	☐	☐	☐	☐
8	Provided a good cross section of general medicine	☐	☐	☐	☐
9	Was an experience I would like to have again	☐	☐	☐	☐

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

For PGY2s and PGY3s Only

If you've completed this OSCE before…

					Comments
How well did you do on this OSCE compared with the previous one(s)?	I did much worse on this OSCE ☐	About the same ☐	I did much better on this OSCE ☐	*Not applicable* ☐	

We welcome any comments, feedback, and suggestions—about the OSCE, the Standardized Patients, the whole process…

Appendix L. Survey: Rater Evaluation of OSCE Experience (designed for General Internal Medicine observers in a residency OSCE, usually faculty observer)

Please indicate your response to each of the questions listed in the columns:

		How difficult was this case for the residents?			How much will residents learn from this case?			How well did the SP portray this case?			Overall, how well did the residents (as a group) perform in this case?				
Case #	Case	Too easy	Just right	Too hard	Nothing	Some	A lot	Not well	O.K.	Very well	Poor	Fair	Good	Excellent	What was most surprising about the residents' performance?
		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

Please indicate how much you agree or disagree with each of the items below:

In general, this OSCE…		Strongly disagree	Somewhat disagree	Somewhat agree	Strongly agree
1	Helped residents identify their strengths and areas they need to improve	☐	☐	☐	☐
2	Stimulated residents to learn more about some of the topics covered	☐	☐	☐	☐
3	Taught residents something new	☐	☐	☐	☐
4	Provided residents with valuable feedback	☐	☐	☐	☐
5	Provided me with new information about residents' performance level	☐	☐	☐	☐
6	Gave me some new ideas for teaching	☐	☐	☐	☐
7	Was a lot like real-life clinical encounters	☐	☐	☐	☐
8	Evaluated residents' skills fairly	☐	☐	☐	☐
9	Was enjoyable	☐	☐	☐	☐
10	Provided a good cross section of general medicine	☐	☐	☐	☐
11	Was an experience I, as faculty, would like to have again	☐	☐	☐	☐

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

We welcome comments, feedback, and suggestions—about the OSCE, the rating form, the SPs, the whole process….

Comments:

Appendix M. Survey: SP Evaluation of OSCE Experience

Please indicate your response to each of the questions listed in the columns:

		How difficult was this case for the participants?			How much will participants learn from this case?			Overall, how well did the participants (as a group) perform in this case?				What was most surprising about the participants' performance?
Case #	**Case**	**Too easy**	**Just right**	**Too hard**	**Nothing**	**Some**	**A lot**	**Poor**	**Fair**	**Good**	**Excellent**	
		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

Were there any consistent problems with using the checklist? Items that didn't seem to work? Aspects of the participants' performance that weren't reflected?

Was there anything about your character that didn't work? Were there any questions asked that you felt unprepared for?

Did any of the participants say or do anything that you felt unprepared to handle?

How did it go giving feedback? Any problems? Any highlights? Suggestions for improvement?

Is there anything that we can do to better prepare you? (Portraying the case, rating performance, giving feedback, staying alert?)

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Appendix N. Sample USP Case

Included below are SP instructions for the "Trouble Breathing/Asthma" USP case, designed to evaluate resident physician performance and the clinical microsystem in a primary care outpatient setting. Highlighted case details are varied each visit to minimize USP detectability.

SP Instructions: Trouble Breathing/Asthma

Scenario	You are a *23–26-year-old* female with a history of asthma. You've been living in New York City a little over a year.
	You came into the clinic today because your asthma has been much worse in the last week. This past week you've been up a lot at night because of trouble breathing and twice had to use your inhaler three times in one night. You knew this was bad, that you shouldn't be using it so much.
	You've had asthma since you were a little kid, but it's never been that bad. Usually, it only affected you when you had a cold, and it never stopped you from doing anything. When you had a cold, you would cough a lot, but it would get better with an inhaler.
	When you moved from *Baltimore* last winter, your asthma started getting much worse. Over the summer it got better, but this winter it's been bad again. If asked if you think the weather is the reason for your increased asthma, you shrug it off with "I don't know, I just know it's gotten worse."
	You never went to the ER until last winter when you had a bad attack. This winter, you've been to the ER (New York Downtown Hospital) a total of three times over the past winter.
	For the past few months, you have been experiencing a very bad cough (hacking, no phlegm). You also have wheezing, shortness of breath (dyspnea), and chest tightness. This happens especially at night, when you go out into the cold or when you walk upstairs.
	Last week, you had a cold (stuffy, runny nose but no fever) that got better on its own. However, over the last few days your breathing has been much worse. Other ways to describe asthma: "chest tightness—when I take a deep breath, my chest hits a wall halfway through what would be a normal breath."
	You use an albuterol inhaler which helps your symptoms and another inhaler (it is kind of coral colored—you don't remember the name of it [Flovent]) that you are supposed to take every day (if specifically asked, you are supposed to take it twice daily, once in the morning, once at night) but stopped using after two days because "it doesn't do anything." You were prescribed the coral-colored inhaler at the ER when your asthma first got really bad when you moved to NYC last winter. If asked if you still have it, you do—it's somewhere in your bathroom. *You are not aware that the Flovent is a preventive medication.*
	This winter, you've found yourself having to use the albuterol inhaler more than usual (until last year you only used it for rare attacks), about three times a week. Over the past month, you started using the inhaler once or twice a day. It seems like you need to use the inhaler "every time you do anything," including walking up the subway steps, and light housework. Last week, things got even worse and you needed the inhaler three times in one night on two different nights. You don't really like taking the albuterol because even though it helps you breathe better; it makes you anxious and jittery.

(continued)

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

History of present illness	Chief complaint:	Difficulty breathing and asthma attacks
	Where	General respiratory
	When	Problem has been getting increasingly worse in past 3 months
	Quality	Debilitating
	Quantity	Three attacks in one night at its worst
	Aggravating/alleviating factors	Aggravating: cold, nighttime, activity, poor air quality; Alleviating (temporary): albuterol inhaler
	Associated symptoms	Hacking cough, wheezing, shortness of breath, low energy
	Beliefs	You don't like to consider yourself "sick," and prefer not to take medications or see a doctor at all. However, your worry about your job makes you feel like you have to get this "taken care of."
		You have not had a regular doctor since moving from home.
Current life situation	You live with your boyfriend from home in an apartment in *Manhattan (Stuyvesant Town)* that belongs to your grandmother, who can no longer live on her own and is now living with your family back in *Baltimore*. You lived at home through college and moved to NYC with your boyfriend after graduation.	
	Your mom is pretty high-strung and she is getting upset that you are sick all the time. You talk on the phone a lot and she is worried you are missing too much work.	
	You work in a *restaurant/retail store*. You are worried about getting fired because you were home sick a few times over the winter, starting when the weather got cold.	
Personality	You are a quiet and friendly person but a bit intimidated by health care providers.	
Past medical history	Besides your asthma, you've had no medical problems. Never been hospitalized. You have had all your vaccinations (your mother has the "little yellow book" where these are written down).	
Family medical history	Your parents have no medical problems. You are not aware of anyone else in your family having asthma.	
Medications	Regular albuterol inhaler, plus "coral-colored" inhaler prescribed at the ER. No previous medications prescribed for asthma. You have been on birth control (Yaz) for 4 years.	
Allergies	You don't have any allergies to medicines. Cats usually bother your asthma; you've never had pets. Cigarette smoke also makes you cough.	
Social history	**Sexual history**	You've been with your current boyfriend since you were both seniors in high school. You had two sexual partners before him, and you always use condoms.
	Smoking	You've never smoked, and no one at home or at work smokes.
	Alcohol/drugs	Occasionally you have a beer. No drugs.
	Nutrition	You eat mostly healthy food. No recent weight gain.
	Exercise	No exercise besides being on your feet all day at work.
Interview challenges for resident	Take a focused history concerning asthma symptoms now and over the past year.	
	Explore patient's motivation for taking medications (stop coughing, keep up at work, stop going to ER).	
	Recommend/counsel on using medications regularly and keeping doctor's appointments.	
Encounter beginning	State how you've been feeling the last few days. If asked about how the problem has been in the past, explain the worsening of the condition this week and last winter and how you've been to the ER a few times.	
Middle	If the resident does not ask about how your asthma is affecting you, state that you are missing a lot of work and sleeping pretty poorly which makes it hard to have any energy. If asked about this, state you're actually pretty worried.	
	If asked about taking medications regularly (not just "when you need them"), you state that you are a little reluctant to do so. You actually don't think the coral-colored inhaler really works since you didn't feel anything when you used it. You are worried about using an inhaler in front of your boss or coworkers because you feel like they will think you are weak and sickly, but it is also pretty embarrassing that you can't run up the stairs without huffing and puffing. You have never seen anyone else use an asthma pump. Sometimes you are not sure whether you are using the pump correctly (take out your pump at that time to give the physician an opportunity to let you demonstrate how you do it). *(We will show you how to do it a little wrong.)*	
	If medications are explained and your understanding of them is checked, state that you are willing to take the two daily preventive pumps of Flovent. You are motivated to get better because you feel horrible and hate going to the ER. You want to be "normal." You would be willing to see a doctor regularly if you didn't have to miss work.	
	If the resident does not come up with a follow-up plan or medication plan, say something like "maybe I'm on the wrong medications…"	
End	You're pleased about having received more information about your problem. You are an intelligent person, and no one had ever explained to you before that the two asthma meds worked in different ways. You are happy about the prospect of getting your asthma under control.	

Appendix O. Sample USP Rating Form

The following resident performance and clinical microsystem checklist ("Trouble Breathing/Asthma" case, Appendix N) is completed by the SP after the USP encounter.

Date:	
MD Name:	

Asthma

SP Name: _____

Clinic Team _____

When did you arrive at your appointed clinic area? _____:_____ am/pm

When did your visit with the resident begin? _____:_____ am/pm

When did your visit end? _____:_____ am/pm

Comments

The primary care associate…			***Comments***		
1st PCA	**2nd PCA** (if applicable)		**1st PCA**	**2nd PCA**	
☐	☐	Greeted me within a reasonable time frame			
☐	☐	Introduced self			
☐	☐	Wore a visible name tag			
☐	☐	Asked me my name			
☐	☐	Asked me my date of birth			
☐	☐	Washed hands before touching me			
☐	☐	Measured my height			
☐	☐	Took my blood pressure			
☐	☐	Weighed me			
☐	☐	Screened for depression using the PHQ-2			
Acknowledged/apologized for any delays		**1st PCA**	☐ No	☐ Yes	☐ NA (no delays)
		2nd PCA	☐ No	☐ Yes	☐ NA (No Delays)
Was friendly and/or professional		**1st PCA**	☐ Rude	☐ Professional	☐ Friendly
		2nd PCA	☐ Rude	☐ Professional	☐ Friendly
Took care to explain things to me		**1st PCA**	☐ No explaining	☐ Some explaining	☐ Fully explained
		2nd PCA	☐ No explaining	☐ Some explaining	☐ Fully explained
Overall, were you satisfied overall with the way the PCAs treated you?			☐ Not satisfied	☐ Somewhat satisfied	☐ Very satisfied
Comments:					

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Experience with clinic			
It was easy to navigate through the system	☐ Not so easy	☐ Relatively Easy	☐ Very easy
The **team** to which I was assigned functioned well	☐ Problems	☐ Functioned O.K.	☐ Functioned Well
Overall, I was treated professionally by non-MD staff	☐ Not at all professional	☐ Somewhat professional	☐ Very professional
Comments:			

Communication Skills

Information Gathering	**Not done**	**Partially done**	**Well done**	*Comments*
Elicited your responses using **appropriate questions**	Asked leading questions AND more than one question at a time	Used leading questions OR asked more than one question at a time	Asked questions one at a time without leading you in your response	
Managed the **narrative flow** of your story	Not able to elicit your story because questions not organized logically	Elicited main elements of story, but illogical order of questions disrupted flow	Elicited full story by asking questions that facilitated natural flow of story	
Clarified information throughout by repeating to make sure understood you	Did not clarify (did not repeat back to you the information you provided)	Repeated info you provided but did not give you a chance to indicate accuracy	Repeated info and directly invited you to indicate accuracy	
Allowed you to talk **without interrupting**	Interrupted	Did not interrupt BUT cut responses short, not enough time	Did not interrupt; allowed to express thoughts fully	

Relationship Develop	**Not done**	**Partially done**	**Well done**	*Comments*
Communicated concern or intention to help	Did not communicate intention to help/concern	Words OR actions conveyed intention to help/concern	Actions AND words conveyed intention to help/concern	
Nonverbal behavior enriched communication (e.g., eye contact, posture)	Nonverbal behavior was negative OR interfered with communication	Nonverbal behavior demonstrated attentiveness	Nonverbal behavior facilitated effective communication	
Acknowledged emotions/feelings appropriately	DID NOT acknowledge emotions/feelings	Acknowledged emotions/ feelings	Acknowledged and responded in ways that made you feel better	
Was **accepting/nonjudgmental**	Made judgmental comments OR facial expressions	Did not express judgment but did not demonstrate respect	Made comments and expressions that demonstrated respect	
Used words you understood and/or explained **jargon**	Consistently used jargon WITHOUT further explanation	Sometimes used jargon AND did not explain it	Explained jargon when used OR avoided completely	

Educ and counseling	**Not done**	**Partially done**	**Well done**	*Comments*
Asked questions to see what you understood (checked your understanding)	Did not check for understanding	Asked if patient had any questions BUT did not check for understanding	Assessed understanding by checking in throughout	
Provided **clear explanations**/ information	Gave confusing/no explanations—made it impossible to understand	Info was somewhat clear BUT still led to some difficulty in understanding	Provided small bits of info AND summarized to make sure clear	
Collaborated with you in identifying possible next steps/plan	Told patient next steps/ plan (OR no next steps/ plan)	Told patient next steps THEN asked patient's views	Discussed options THEN mutually developed plan	

Org and time man	Not done	Partially done	Well done	*Comments*
Managed time effectively	Paced the encounter poorly; did not manage time well	Paced the encounter, managed time to cover most of what needed to be covered	Paced the encounter very well; managed time so that visit seemed complete	

Resident Case-Specific Skills

Assessing history	Not done	Partially done	Well done	*Comments*
Asked for **name and date of birth**	Did not ask	Asked for either name only or date of birth only	Asked for name and date of birth	
Asked about **past medical problems**	Did not ask patient past medical problems	Asked if patient has any past medical problems but not specific	Asked a comprehensive past medical history—including meds, allergies	
Asked about **alcohol use**	Did not ask	Asked about BUT NOT quantity or frequency	Asked about AND assessed quant. and freq.	
Asked about **drug use**	Did not ask	Asked about BUT NOT quantity or frequency	Asked about AND assessed quant. and freq.	
Asked about **smoking**	Did not ask	Asked about BUT NOT quantity or frequency	Asked about AND assessed quantity and frequency	
Asked about **work history** and **educational level**	Did not ask	Asked about current job but not work history and/or educational level	Asked about all	
Asked about **social and family support**	Did not ask	Asked questions about family/friends	Identified access to support	
Asked about **family medical history**	Did not ask	Asked generally but not specifically	Obtained a full family medical history	
Asked about **depression**	Did not ask	Asked generally about depression but did not use the PHQ-2	Asked about depression using at least the PHQ-2 (asked about **lack of interest** AND **mood**)	
Offered **HIV** screening	Did not offer	Offered test (learned that you are HIV-negative)		
Asked about **tetanus** and **other immunizations**	Did not ask	Asked about one vaccine	Asked about more than one vaccine	

Yes	**Not sure**	***Review of systems*** Asked about…	**Yes**	**Not sure**	***Physical exam***
☐	Not sure	**Skin** rash, itching, pigmentation, dryness; hair growth or loss	☐ PCA ☐ Intern	Not sure	**Vital signs** measured blood pressure, took pulse
			☐	Not sure	**Washed hands before exam**
☐	Not sure	**Eyes/ears/nose/mouth/throat** vision, hearing, throat pain, headache	☐ *Please circle*	Not sure	**Eyes/ears/nose/mouth/throat** inspected
☐	Not sure	**Cardiovascular** chest pain, palpitations, shortness of breath, walking	☐	Not sure	**Heart** listened
			☐	Not sure	**Checked extremities** felt pulses, inspected hands/feet
☐	Not sure	**Respiratory** shortness of breath, wheezing	☐	Not sure	**Lungs** listened, palpated, and/or percussed
☐	Not sure	**Musculoskeletal** pain, swelling, redness/heat muscles/joint; range of motion	☐	Not sure	**Strength/range of motion** inspected and tested muscles and joints
Yes	**Not sure**	***Review of systems*** Asked about…	**Yes**	**Not sure**	***Physical exam***
☐	Not sure	**OB/GYN** pregnancy, menstruation, last pap smear, gyn health	☐	Not sure	**Abdomen** inspected, listened, palpated, and/or percussed
☐	Not sure	**Gastrointestinal** bowel movements, pain, swallowing, appetite	☐	Not sure	***Other*** _____
☐	Not sure	**Allergic/immunologic/lymphatic/endocrine** reactions to drugs, food, insects, skin rashes; trouble breathing; anemia; lymph nodes	☐	Not sure	***Other*** _____

Patient education	**Not done**	**Partially done**	**Well done**	*Comments*
Assessed your **understanding of asthma**	Did not assess	Obtained a full history of your personal experience of asthma OR asked what you know about the condition	Fully explored both your personal experiences of asthma and understanding of the condition	
Assessed your **understanding of asthma medications**	Did not asses understanding	Told you how asthma medications work without assessing understanding	Assessed understanding and corrected misinformation	
Checked/demonstrated **inhaler technique**	Didn't address inhaler technique	Demonstrated or explained correct use but didn't check inhaler technique	Checked your technique and demonstrated correct use	
Recommended that you use the **controller/preventive inhaler (Flovent)** daily for better symptom management	Didn't recommend	Suggested that you should use the Flovent inhaler daily	Gave a clear and direct recommendation that you should use the Flovent and explained how it would better manage symptoms	
Recommended short course of **prednisone (oral steroids)**	Did not recommend	Recommended several days of steroids		
Recommended that you use your **inhaler with a spacer**	Did not recommend	Recommended that you use spacer		
Gave list of prescribed medications	Did not give list of medications	Gave list but did not discuss medications prescribed	Gave list and fully explained medications prescribed	
Labs/referrals ordered ___ Respiratory therapy referral ___*Other*: _____	Ordered no labs	Offered labs/referrals but did not explain which ones and rationale behind decision	Offered labs, explained choice and rationale for labs and discusses follow up of results	

Made which of the following **health maintenance** recommendations:	**Yes**	**NS**	**Health maintenance recommendation**	*Comments*
	☐	Not sure	Take preventive medication (Flovent)	
	☐	Not sure	*Other*: _____	

Counseling (Management & Treatment)	**Not done**	**Partially done**	**Well done**	*Comments*
Reviewed **plan** with you	Did not review summation of visit and plan	Reviewed plan but did not assess ability/willingness to comply	Reviewed plan, assessed ability, willingness to comply	
Asked you what **further questions** you have	Did not ask	Asked about questions but in a brisk manner, didn't allow sufficient time	Asked you what further questions you had in a way that invited questions	
Gave information about **follow up and further contact** w/ MD	Did not address	Addressed follow-up but was not specific	Specifically addressed follow-up, setting time and person	
Helped you understand how to **navigate** the system in order to **follow through** on next steps	Did not help navigate the system	Partially explained how system works in terms of next steps (blood work, etc.)	Fully explained process and how to navigate the system	

Patient-Centeredness/Satisfaction

The resident …	Not done	Partially done	Well done	*Comments*
Answered or addressed all my **questions**/concerns	Only answered/addressed a few of the most central	Answered/addressed many of my questions/concerns	Answered/addressed all of my questions/concerns	
Took a **personal interest** in me; treated me as a **person**	Did not see me as a person	Viewed me as a person, but did not take personal interest	Took an active personal interest in me	
Gave me **enough information**	Not given much info at all	I was given some information, but I still had questions	I was given all the information I wanted/needed	
Made you feel like had **enough time** (not rushed)	Did not have enough time; visit felt rushed	Mostly had enough time (visit a bit rushed); felt some time pressure	Felt no real-time pressures; covered most w/out pressure	

Activating the Patient

This encounter….	Not done	Partially done	Well done	Comments
This encounter helped me to understand the nature and causes of asthma	Did not help me understand	Helped me understand some things but not everything	Helped me fully understand what happened	
After the encounter, **I understood how to manage my asthma in the future** (including how medications work and how to use them)	I did not learn about asthma management	I found out about some of the treatment options	I left with a clear treatment and management plan	
This visit made me feel **confident I can keep asthma interfering** too much with my life	Did not affect my confidence	Helped me feel more confident that I could keep asthma w/ life	Helped me feel very confident that I could keep asthma from interfering w/ my life	
Because of this encounter, I am **confident I can figure out solutions** if something new comes up	Did not affect my confidence	Helped me feel somewhat confident that I could deal with new issues	Helped me feel quite confident that I could deal with new issues	

Overall Recommendations

Would you recommend this doctor to a friend or family member for his/her…

Communication/interpersonal skills?	Not recommend	Recommend with reservations	Recommend	Highly recommend
Medical competence? Application of medical knowledge	Not recommend	Recommend with reservations	Recommend	Highly recommend
Professionalism? Accountable, respectful, sensitive and/or responsive, compassionate	Not recommend	Recommend with reservations	Recommend	Highly recommend

Would you recommend this **clinic** to a friend if they needed primary care?

Not recommend	Recommend with reservations	Recommend	Highly recommend
Please explain your choice (comment on anything you feel is relevant including the facility, staff, waiting area, and time spent waiting):			

Detection

Do you think this physician recognized that you were a standardized patient?

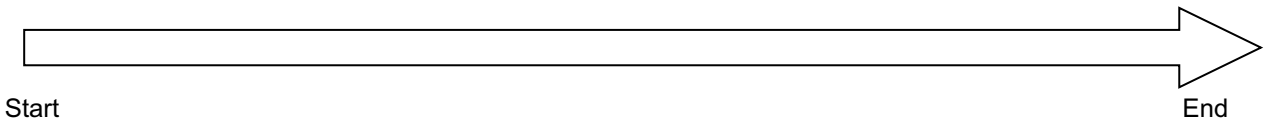
No	Yes	*If yes, explain why*

What materials did you receive during this visit?

Yes	Materials
☐	Lab orders
☐	Health education pamphlets/information
☐	Contact information
☐	Follow-up appointment slip
☐	Handwritten note, diagram, explanation
☐	Spacer or inhaler
☐	*Other*: _____

Please use the following timeline to depict the sequence and timing of the visit.

Divide the timeline into the four major segments of the case: history gathering (HG), physical examination (PE), counseling about asthma and medication (ASTH), report to preceptor (PRE), and health recommendations (HR). Place them in the same order as in the visit, and do your best to represent the portion of the visit that was spent on each. If time during the visit was spent on other issues, please describe and put on the timeline too!



Appendix P. Other Resources for OSCE Development

Station Development

- Search MEDLINE (http://www.ncbi.nlm.nih.gov/pubmed) and the Internet for articles describing OSCEs and OSCE stations.
- MedEdPortal (www.mededportal.org) provides a peer-reviewed collection of free educational resources including cases and OSCE stations.
- OSCE exam preparation books (e.g., Hurley, 2005) and websites (e.g., OSCE Home: www.oscehome.com/) contain station examples.
- Consider non-OSCE Clinical Vignettes that can be converted into OSCE cases.
- The Association for Standardized Patient Educators (ASPE: www.aspeducators.org) includes a virtual library with resources for station development (some resources require membership in the organization).
- Professional listservs/blogs may require registration but can provide opportunities to access expertise and resources worldwide: SP-Trainer (mailman2.u.washington.edu/mailman/listinfo/sp-trainer), the official ASPE listserv; DR ED (list.msu.edu/cgi-bin/wa?A0=dr-ed), an international general listserv focusing on medical education; and eGroups of the Society for Simulation in Healthcare (www.ssih.org).

Standardized Patient Recruitment and Training

- The Association for Standardized Patient Educators (ASPE) holds annual conferences and gives out annual project awards which provide further resources (e.g., feedback training, recruitment, and training of multicultural SPs). Their website (www.aspeducators.org) includes a searchable bibliography organized in the following sections: Overviews of SP Use, Project or Program Evaluation, SPs in Teaching Exercises, OSCEs, Measurement Tools, Influence of being an SP on the SP/Special Populations of SPs, and Models and Computers for Simulation.
- Wallace (2007) provides an excellent in-depth resource for SP coaching.

Educational Research and Psychometrics

- Look for university courses on educational measurement in departments of education or psychology.
- The Association of American Medical Colleges (AAMC) runs a number of relevant training programs, such as the Medical Education Research Certificate (MERC) (www.aamc.org/members/gea/merc) through their Group on Educational Affairs (GEA).
- Consult the Foundation for Advancement of International Medical Education and Research (FAIMER: www.faimer.org) for fellowship opportunities.

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Appendix Q. PICOT Worksheet for Use When Designing an SP-Based Research Question

Study:		**Title:**
Research question:		
Population	***Target group/sample of learners***	
Intervention/ independent variable	***"Treatment" (curricular intervention, innovation or variable expected to affect outcome)***	
Comparison	***Reference group (counterfactual—what happens without Intervention?)***	
Outcome	***Measure reflecting effectiveness of intervention***	
Time	***Duration for study; when to measure outcome***	

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Appendix R. Sample Telemedicine Assessment Items to Incorporate in SP Rating Form for a Virtual Telehealth OSCE

The following is the domain of telemedicine assessment items that we add to the SP checklist for virtual telemedicine OSCEs. This domain can be incorporated into a full checklist (like the example found in Appendix H) so that regular communication domains and case-specific skills are assessed in addition to telemedicine skills.

Telemedicine skills	Not done	Partially done	Well done	*Comments*
Partnered with you to perform a **virtual physical exam**	Made no attempt to perform physical exam	Asked patient to perform basic exam maneuvers or utilize peripheral monitoring devices (thermometer, home BP cuff, Fitbit/Apple Watch, etc.)	Asked patient to perform maneuvers or access peripheral monitoring device FOLLOWED BY verbal confirmation of findings with patient or collateral	
Confirmed patient identifiers	Did not attempt to identify patient or patient's location	Asked patient to confirm either (a) name and/or date of birth (b) call back number OR (c) location	Asked patient to confirm (a) name and/or date of birth (b) call back number AND (c) location	
Exhibited **comfort** and **confidence** using video interface	Was shy/uncomfortable in front of the camera, let technological glitches (if any) distract from the encounter	Mostly comfortable on camera, occasional stumbling but interview was not derailed by occasional video delay or glitch	Confident on camera, acknowledged and moved forward from technical glitches, and did not let video interface detract from natural conversation	
Used **non-verbal communication** to enrich communication on camera	Avoided eye contact via the webcam, slouched, or was angled away or too far from camera	Made occasional eye contact with webcam, sometimes slouched or out of view	Maintained eye contact with webcam throughout encounter, sat squarely in front of camera, and at appropriate distance	
Utilized live video to **augment information gathering**	Made no attempt to visually reconcile medications, witness reproducible symptoms, talk with onsite collateral (family/HHA/VNS), visual tour of home	Did one of the following: visually reconcile medications, witness reproducible symptoms, talk with onsite collateral (family/HHA/VNS), visual tour of home	Did two or more of the following: visually reconcile medications, witness reproducible symptoms, talk with onsite collateral (family/HHA/VNS), visual tour of home	
Actively optimized technical aspects of the virtual encounter	Did not assess sound (e.g., volume, clarity, background noise), video (e.g., pixilation or delay), or "backup plan" if technology failed (e.g., phone call)	Assessed two of the following: sound (e.g., volume, clarity, background noise), video (e.g., pixelation or delay), or 'backup plan' if technology failed (e.g., phone call)	Assessed three of the following: sound (e.g., volume, clarity, background noise), video (e.g., pixilation or delay), AND "backup plan" if technology failed (e.g., phone call)	
Maintained appropriate **computer etiquette** during encounter	Frequent typing without explanation, appears preoccupied with computer, or distracted	Occasional types with minimal explanation of actions	Paused video or provided clear explanation while documenting, searching another website, or having another screen open for the purpose of patient care	

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Appendix S. Example OSTE Checklist

OSTE: Student Feedback
Faculty Coded ID:

	Not done	Partly done	Well done	*Comments*
Feedback skills				
Clarify goals and objectives for the feedback	Did not clarify goals and objectives	Somewhat clarified goals and objectives	Clarified goals and objectives for feedback	
Asked learner to evaluate his own performance	Learner not asked to evaluate his own performance	Awkwardly asked learner to evaluate his own performance	Learner skillfully asked to evaluate his own performance	
Responded to and built upon learner's self-assessment	Did not respond to or build upon learner's self-assessment	Made an attempt to respond to or build upon learner's self-assessment	Effectively responded to and built on learner's self-assessment	
Built on positive behavior then moved onto behavior that could be improved	Did not build on positive behavior and did not move from positive behavior to behavior that could be improved	Made an attempt to build on positive behavior and move onto behavior that could be improved	Appropriately and skillfully built on positive behavior and them moved onto behavior that could be improved	
Focused on specific behaviors	Did not focus on specific behavior	Made an attempt to focus on specific behaviors	Effectively focused on specific behaviors	
Used nonjudgmental language	Used judgmental language	On occasion, used language that could be interpreted as judgmental but corrected it	Used nonjudgmental language	
When feedback was subjective, labeled it as such	Did not label subjective feedback appropriately	On occasion, did not label subjective feedback as such, but usually labeled it correctly upfront	Appropriately labeled subjective feedback	
Acknowledged and addressed emotions during feedback session	Did not acknowledge or address emotions during session	Made an attempt to acknowledge and address emotions during session, but was awkward at times	Skillfully acknowledged and addressed emotions during feedback session	
Problem-solved together on ways to change behavior	Did not problem-solve with learner	Attempted to problem-solve with learner but could have been done more skillfully	Effectively problem-solved with learner	
Learning climate				
Actively listened to learner	Did not appear to listen to learner	Listened to the learner somewhat	Appeared to actively listen to learner (looked at learner, nodded head)	
Asked effective questions to engage learner	Did not ask questions of learner	Engaged leaner somewhat. Asked few questions	Asked effective questions that engaged learner	

(continued)

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

	Not done	Partly done	Well done	*Comments*
Encouraged learner to express concerns and needs	Did not encourage learner to bring up concerns or needs	Made some attempt to encourage learner to bring up concerns or needs but could have been more effective	Made learner feel comfortable bringing up concerns and needs	
Expressed respect for learner's experience and emotions	Treated learner in an insensitive manner. Discouraged learner from discussing his experience or emotions	Treated learner in a neutral manner, neither respectful or disrespectful	Encouraged learner to discuss experiences, made learner feel comfortable discussing emotions	
Expresses empathy and/or concern for learner	Did not express empathy or concern for learner	Expressed some empathy and/or concern but not in a way that made you feel better	Sincere expressed empathy and/or concern for learner	
Close session				
Summarized the session	Did not summarize the session	Attempted to summarize the session but left out important information	Effectively summarized the session	
Ended with a positive comment	Did not end with a positive comment	Ended with a positive comment but it was somewhat awkward	Ended with a positive comment	
Plan for the future	Did not make a plan for the future with the learner	Made a plan for the future but it was incomplete, unrealistic, awkward	Effectively made a plan for the future with the learner	

Overall, how would you rate the overall performance of this teacher?

1	2	3	4	5
Inadequate (Ineffective likely to create learning problems; e.g., confusion or dissatisfaction)	**Marginal** (Uses some skills effectively but others may create learning problems)	**Competent/adequate** (Uses most skills effectively)	**Very effective** (Uses all skills effectively, minor suggestions would enrich)	**Exemplary** (At the level of an experienced teacher whose skills are so good you would seek out learning opportunities)

Comments:

Appendix T. Example Standardized Learner Instructions: Medical Student Feedback Case

This checklist was developed by Linda Tewksbury, MD, Cynthia Osman, MD, and Lucy Chang, MD.

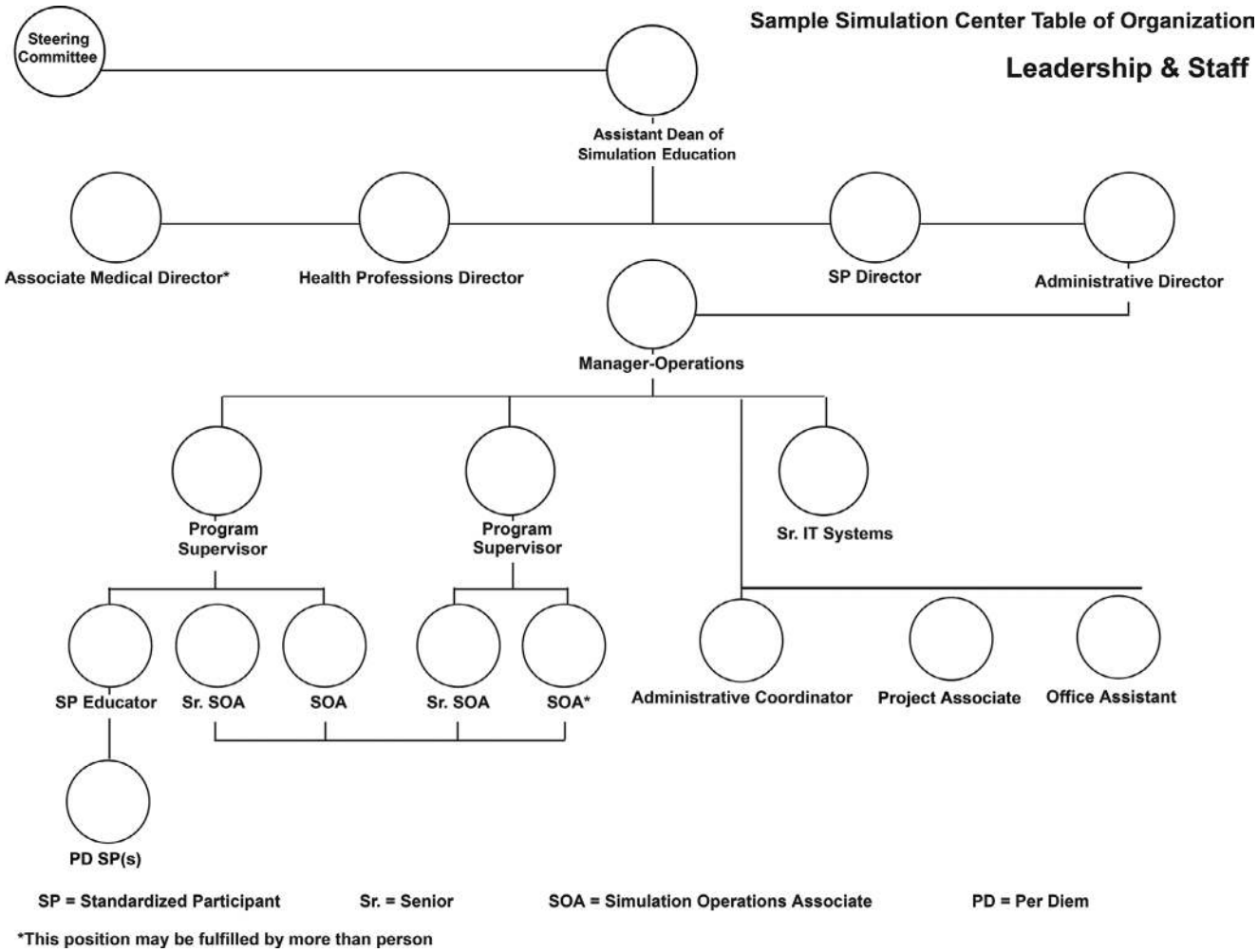
The scenario	Your name is Susan Mills. You are a second-year medical student on your pediatric clerkship clinical rotation. After rotating for 2 weeks in the pediatric clinics, you have just finished your first of 2 weeks on one of the inpatient units. The inpatient team includes the ***ward attending*** (senior doctor on the pediatric faculty who oversees caring for the patients on the ward and teaching), ***senior resident*** (doctor in second year of training) two ***interns*** (doctor in first year of training), and two other students. It is "feedback Friday," and you are really looking forward to getting some feedback on your performance from your ward attending. It has been a bit of a difficult adjustment as this is only your second rotation, but you have been working very hard and pleased that you really are enjoying working with the kids.
Feedback challenge for faculty	• To give you some negative feedback regarding your professionalism without making you defensive • To give you constructive suggestions on how best to address your areas of weakness, including the professionalism issues
Personality/ context	You were very excited to start pediatrics as it is a specialty you are considering, but you were nervous that you would not be able to handle the sicker children. You were a camp counselor in high school and volunteered at a preschool but have never been around very sick children before. Your first rotation was in Radiology so you did not get that much interaction with patients. You have been pleased to find that you do enjoy working with sick children and have become particularly attached to a 3-year-old admitted with cellulitis (Michelle) and a troubled teenager admitted after a motor vehicle accident (John). You feel like you have been working very hard but are still trying to figure out how to balance all the expectations of the clerkship (studying for shelf exam vs. preparing for presentation vs. preparing write-ups, etc.).
Your experience on the pediatric clerkship to date	As this is your first "major" clinical rotation, it has been hard to keep up with all the abbreviations that are used on rounds and in conferences. It has been very helpful to be able to look up things on your phone so you can follow the discussion.
	You were assigned to the pediatric clinics for the first 2 weeks of the rotation, which you enjoyed, but it was hard getting used to going to a different clinic every day. You were eager to start on the inpatient unit this past week, but it has also been hard as you feel you are just starting to get the hang of things and now there is only 1 week left. The two students who are working on the team with you, Jeff and Nina, are nice enough but are always jumping in to answer questions, even when the question is directed at you. While you do not think of yourself as a "shy" person, you tend to be on the quieter side and certainly don't want to be the kind of student who tries to overshadow her classmates.
	You were happy to be given the opportunity to prepare a brief presentation on cellulitis (a bacterial infection of the skin), but you were more nervous than expected. As you were asked to give only a 5-minute presentation, you tried to just stick to the basic facts and were surprised at some of the more detailed questions you were asked. The faculty had several questions about "MRSA" (pronounced "mersa" which stands for methicillin-resistant *Staph. aureus* infection—a particularly serious cause of cellulitis). You did not spend time researching this and had difficulty answering questions about changes in prevalence of "community acquired" vs. "nosocomial" MRSA infections and how best to treat.
	This week was also challenging as you had a presentation for your ***preceptor*** this past Wednesday (your ***preceptor*** is a doctor on the pediatric faculty who meets with you and a small group of students once a week throughout your 6-week rotation to go over cases with you. This person is different than the ***ward attending*** from whom you are now getting feedback). You were happy that your intern allowed you to go the library on Tuesday afternoon since the ward was quiet, to work on your presentation. You came back after "sign out" to check on your patients, but your team had already left for the day.
	To add to your stress, you are a little bit more distracted than usual as your mother is having elective surgery next week (having a "lipoma"– a benign fat tumor—removed from her back). Although you are not particularly worried about the surgery as it is pretty routine, your mother has a million questions and has been texting you all week and wanting to talk at night.

(continued)

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

If asked to self-assess your performance:	If first asked how you think it has been going on the inpatient unit before given feedback, you immediately state that "it's going well!" You have been enjoying the rotation, and you feel like your presentations on rounds in the morning and your notes have been good, especially after getting some feedback from your intern on the first day. If asked what you think is your strength, you proudly talk about the two patients you have really connected with. There is a 3-year-old girl, Michelle, who was admitted for cellulitis whose mother is always at the bedside. Although the patient initially would cry when anyone walked into the room, you have made a great effort to try to spend time with her, and she has finally come to trust you and happily lets you examine her. Her mother also seems to trust you and welcome your visits. You have also been assigned to a 16-year-old boy, John, from troubled family (single mom with other young kids and father a drug addict currently in prison) who was recently in a motor vehicle accident. He has hardly had a visitor and has been very sullen all week, but you finally got him talking about what his life is like at home and his father who was verbally abusive before going to jail.
	If asked for what you think you could improve on, you state you know you could have done a better job on your cellulitis presentation and probably should have asked for more clarification as to what was expected. You realize that you might not be using your study time as effectively as you should (for example, you should have put more time into your cellulitis presentation instead of working most of the night on your preceptor presentation).
Reactions to feedback:	You have been looking forward to getting feedback today from your attending. You are very, very surprised and disappointed to learn that the residents think you are "disinterested." While your personality isn't bubbling with enthusiasm, you really are interested in pediatrics and want to learn as much as possible. You are very surprised to get criticized for being on your phone during conference as "everyone" seems to do that and it has been helpful to you. The only time you took out your phone on rounds was when you noticed and "urgent" text from your mom. You do understand, when pointed out to you, how it could be misperceived as not interested and certainly will be mindful not to do so in the future.
	You are shocked if you are criticized for "disappearing" on Tuesday afternoon. You realize that your intern had clinic that afternoon and probably did not tell the senior resident she had given you permission to leave.
	If the attending interacts with you respectfully and starts by getting your input before giving feedback, while still a little defensive when you initially hear negative feedback, you really do want to know what you could do better and try to be open, especially to the suggestions. You are especially receptive if the ward attending tries to objectively discuss actions ("we noticed you were on your phone during conference") and perceptions ("it seemed to us that you were not following the discussion") rather than more subjective labeling ("you were rude and disrespectful").
	If the attending does not give you a chance to weigh and immediately starts with the negative feedback, you are first shocked and quiet then start to be a little angry and defensive (i.e., when told you "disappeared" on Tuesday, you state "well I told the intern before I left and he said it was OK" or regarding being on your phone "the residents and even attendings are on their phone too!)

Appendix U. Sample Simulation Center Table of Organization: Leadership and Staff



S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Appendix V. Simulation Operations Associate (SOA): Novice SOA Orientation Schedule

Orientee:
Primary Preceptor:
Expected Orientation Period: [X weeks]
Week [#]: [Date]—[Date]

Day 1: [Date] [time scheduled]; [preceptor time scheduled] [remote or at SIM Center]	
Time	**Learning activity**
	New hire modules
	SIM Center weekly team meeting
12p–1p	**Lunch**
	Independent work: Review SOA syllabus Explore SIM Center website Explore organizational website Google/YouTube: what is an OSCE? Write a brief bio for our website
	Debrief with preceptor

Day 2: [Date] [time scheduled]; [preceptor time scheduled] [remote or at SIM Center]	
Time	**Learning activity**
	Check emails and settle in
	General SIM Center orientation: • Overview of SIM Center • Overview of center hours • Organizational staff chart (website) • Key documents/shared drives • Overview of program reservation process • Overview of SIM Center courses • Organizational website overview – Compliance modules – Timekeeping application and kiosk – Protocols • Comprehensive tour • Environmental walk-through – Location of elevator keys – Lockers – Fire emergency exits and alarm codes • Identification cards
12p–1p	**Lunch**
	SP weekly meeting
	Intro meeting with key staff (other SOAs, program supervisors, program managers, operations manager, administrative director, SP program director, dean of simulation) *To be scheduled over the first month at varying times
	Debrief with preceptor

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

Day 3: [Date] [time scheduled]; [preceptor time scheduled] [remote or at SIM Center]	
Time	**Learning activity**
	Introduction to learning management system
	Independent work Hands-on practice with learning management system
	Observe OSCE encounters
12p–1p	**Lunch**
	SP operations weekly meeting
	SIM Center IT orientation
	Debrief with preceptor

Day 4: [Date] [time scheduled]; [preceptor time scheduled] [remote or at SIM Center]	
Time	**Learning activity**
	Check email, settle in
	Overview of SOA role, types of programs, organizations/structure
	Review SPs and SP/OSCE meetings (structure, goals, attendees)
12p–1p	**Lunch**
	In-depth tour of OSCE equipment/storage
	Debrief with preceptor

Appendix W. SP Program Intake and Program Design Checklist (IPDC)

1. **Program title:**
2. **Date received reservation:**
3. **Date of prescreen:**
4. **Date of intake:**
5. **Date of follow-up intake:**
6. **Program previously held:**
7. **Modality of program:**
8. **Location of program:**
9. **Organization:**
10. **Discipline or GME program:**
11. **Number of learners:**
12. **Level of learners:**
13. **External learners:**
14. **Number of faculty:**
15. **Faculty lead/contact info:**
16. **Coordinator/contact info:**
17. **SOA leading intake:**
18. **SIM center director at intake:**

Reservation details:

Date	Time	Rooms

Pre-screen information:

19. **Are you running this program to**: ☐ Assess ☐ Teach
20. **If assessment:** ☐ Summative ☐ Formative
21. **What are the goals and objectives?** (Copied from reservation form if applicable)
    1.
    2.
22. **Case names**
    1.
    2.
    3.

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

23. **Which of the following content applies?**

☐ Communication	☐ Patient education and counseling	☐ Treatment plan and management
☐ Documentation	☐ Physical exam	☐ Interprofessional collaboration
☐ Procedural skills	☐ Telehealth	☐ Professionalism—TransferofCare/iPASS
☐ Professionalism—bad news	☐ Professionalism—difficult situation	☐ Professionalism—informed consent
☐ Professionalism—DEI (diversity, equity, inclusion)		☐ Teaching/precepting

**If hybrid SIM/OSCE, SIM Center to reach out to SIM program supervisor for those content items*

What types of data and reports would you like to obtain?

24. Video recording? ☐ Y ☐ N	**IT ticket:** ☐ Video view access ☐ Video download/keep

- *Insert information about video retention and storage policies, specific to organization*

25. **Reporting:** ☐ Y ☐ N
    - ☐ **Answer analysis report**: Aggregate response histograms and mean scores per question
    - ☐ **Complete report of questions and answers/scores:** Individual learners (PDF) and group (raw data)
    - ☐ **Batch learner reports**: Individual learner zip folder with assessments
    - ☐ **Assign scoring rubric for learner notes.** Specify instructor(s):
    - ☐ **Release scores/assessments in learner portfolio (must be same date as BLINE video access, if applicable)**
26. **Standardized participants:** ☐ Y ☐ N
    - ☐ SP recruitment and training (**notify SP educator and enter details**)

- **SP arrival time (30 min prior to first encounter):**
- **SP departure time (15 min after last encounter):**
- Case name(s) and preferred demographics (age/gender/ethnicity); specify if NEW:
  - Case 1:
  - Case 2:
  - Case 3:
- Number of principal SPs per case:
- Number of back-up SPs (specify if cross trained):
- Total number of SPs:

SP training planning:

☐ Just-in-time training (same date as program) ☐ Training in advance of program date

Location of training: ☐ Remote ☐ SIM Center

- *SP training requires 2–4 hours depending on case/checklist; faculty is required to attend*
- *SP educator will follow up with training options and gather faculty availability*
- *Note: SP rate: $28/hr, 2-hour minimum for in-person assignments*

☐ Referrals: SIM Center provides referrals and program recruits/trains *(insert request form link on SIM Center website)*

☐ Managed by program

If needed:

27. **Remote OSCE details:**

A. Remote OSCE: Location of participants (SP, Learners, Faculty, & Staff):

B. Remote OSCE: Faculty observers: ☐ Y ☐ N

C. Remote OSCE: Staff support on the day of your program: (How many, functions, etc.) ☐ Y ☐ N

Zoom OSCE role	Point of Contact	Responsibilities
Meeting Host	SIM Center	Zoom Meeting Controller, Broadcast banners, Timing
Meeting Co-host	Program Coordinator	Learner Point Person
SP Point of Contact	Program Coordinator	Assessments Recording
Recording (1 per room)	Faculty observers SPs Learners	Must be present in room for entire period of recording Must use desktop or laptop device

Additional notes:

28. **Space requirements:**
Classroom setup: ☐ Horse shoe seating ☐ Classroom seating ☐ Projector
Conference room setup:

29. **Assessment menu:**

Assessment type	Y/N	Prior to OSCE	During encounter	Post-encounter	Off-site
Door note					
Learner post-encounter note					
Assess learner post-encounter note					
SP verbal feedback					
SP assess learner					
Instructor verbal feedback					
Instructor assess learner					
Learner self-assessment					
Learner peer assessment					
Group debrief					
Video review					

30. **Agenda/exam flow:**
•

31. **Program support (control room, hallway monitor, other)**:
•

32. **Action items**:
    SIM Center: **Target due date:**
    1.
    2.
    3.
    Program: **Target due date:**
    1.
    2.
    3.
33. **Program setup:**

Location	Case	Setup
Room 1		
Room 2		
Room 3		

Standard Equipment:

- ☐ Stretcher OR exam bed
- ☐ Plasma screen
- ☐ Desktop computer
- ☐ Phone
- ☐ Blood pressure cuff
- ☐ Sink/soap
- ☐ 2 chairs inside
- ☐ Ophthalmoscope/PanOptic ophthalmoscope
- ☐ Gloves (S, M, L)
- ☐ 1 chair and 2 headsets inner corridor
- ☐ Exam lights
- ☐ Gowns
- ☐ Drapes
- ☐ Additional computers (laptops, computer on wheels)—*insert IT laptop request form needed*

References

Articles, Reports and Books

1. Accreditation Council for Graduate Medical Education (ACGME) Outcomes Project and the American Board of Medical Specialties (ABMS) Joint Initiative. Toolbox of Assessment Methods. ACGME, ABMS; 2000.
2. Accreditation Council for Graduate Medical Education. ACGME Core Competencies. ACGME; 2001. Available online: www.acgme.org/acwebsite/RRC_280/280_corecomp.asp. Accessed April 1, 2012.
3. Aeder L, Altshuler L, Kachur EK, Barrett S, Hilfer A, Koepfer S, Schaeffer H, Shelov SP. The "Culture OSCE" – Introducing a formative assessment into a postgraduate program. Educ Health. 2007;20:11.
4. Adamo G. Simulated and standardized patients in OSCEs: achievements and challenges 1992-2003. Med Teach. 2003;25:262–70.
5. Altshuler L, Kachur EK. A Culture OSCE: teaching residents to bridge different worlds. Acad Med. 2001;76:514.
6. Altshuler L, Kachur E, Krinshpun S, Sullivan D. Genetics Objective Structured Clinical Exams at the Maimonides Infants & Children's Hospital of Brooklyn, New York. Acad Med. 2008;83:1088–93.
7. Anderson MB, Kassebaum DG, editors. Proceedings of the AAMC consensus conference on the use of standardized patients in the teaching and evaluation of clinical skills. Acad Med. 1993;68:437–83.
8. Association of American Medical Colleges (AAMC) Medical School Objectives Project. Report I: learning objectives for medical student education—guidelines for medical schools. AAMC; 1998.
9. Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in Clinical Neurology. J Med Educ. 1964;39:802–5.
10. Boulet JR, De Champlain AF, McKinley DW. Setting defensible performance standards on OSCEs and standardized patient examinations. Med Teach. 2003;25:245–9.
11. Boulet JR, Smee SM, Dillon GF, Gimpel JR. The use of standardized patient assessments for certification and licensure decisions. Simul Healthc. 2009;4:35–42.
12. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. N Engl J Med. 2006;355:2217–25.
13. Carney P, Ward D. Using unannounced standardized patients to assess the HIV preventive practices of family nurse practitioners and family physicians. Nurse Pract. 1998;23(2):56–8, 63, 67-8 passim.
14. Casebeer L, Klapow J, Centor R. An intervention to increase physicians' use of adherence-enhancing strategies in managing hypercholesterolemic patients. Acad Med. 1999;74:1334–9.
15. Chander B, Kule R, Baiocco P, Chokhavatia S, Kotler D, Poles M, Zabar S, Gillespie C, Ark T, Weinshel E. Teaching the competencies: using Objective Structured Clinical encounters for gastroenterology fellows. Clin Gastroenterol Hepatol. 2009;7:509–14.
16. Cleland J, Arnold R, Chesser A. Failing finals is often a surprise for the student but not the teacher: identifying difficulties and supporting students with academic difficulties. Med Teach. 2005;27:504–8.
17. Culver J, Bowen D, Reynolds S, Pinsky LE, Press N, Burke W. Breast cancer risk communication: assessment of primary care physicians by standardized patients. Genet Med. 2009;11:735–41.
18. Cumming AD, Ross MT. Tuning Project (medicine) – learning outcomes / competences for undergraduate medical education in Europe. Edinburgh: The University of Edinburgh; 2008. Available online at: www.tuning-medicine.com. Accessed April 1, 2012.
19. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med. 2003;78:775–80.
20. Downing SA, Tekian A, Yudkowsky R. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. Teach Learn Med. 2006;18:50–7.
21. Epstein R, Franks P, Shields CG, Meldrum SC, Miller KN, Campbell TL, Fiscella K. Patient-centered communication and diagnostic testing. Ann Fam Med. 2005;3:415–21.
22. Glassman P, Luck J, O'Gara E, Peabody J. Using standardized patients to measure quality: evidence from the literature and a prospective study. Jt Comm J Qual Improv. 2000;26:644–53.
23. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J. 1975;1:447–51.
24. Harris G. New for aspiring doctors, the people skills test. The New York Times. 2011 July 10.
25. Hatchett P, Haun C, Goldenhar L. Training standardized patients to give feedback to medical trainees: the state of the art. ASPE Project Awards; 2004. Available online at: www.aspeducators.org/aspe-project-awards.php. Accessed April 1, 2012.
26. Hauer KE, Ciccone A, Henzel TR, Katsufrakis P, Miller SH, Norcross WA, Papadakis MA, Irby DM. Remediation of the deficiencies of physicians across the continuum from medical school to practice: a thematic review of the literature. Acad Med. 2009;84:1822–32.
27. Hodges B. The objective structured clinical examination: a socio-history. Koeln, Germany: LAP Lambert Academic Publishing; 2009.
28. Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. Med Educ. 2003;37(11):1012–6.
29. Hurley KF. OSCE and Clinical Skills Handbook. Toronto, ON: Elsevier Health Sciences; 2005.
30. Institute of Medicine (IOM) Board on Health Care Services Committee on the Health Professions Summit. In: Greiner A, Knebel E, editors. Health professions education: a bridge to quality. National Academies Press; 2003. Available online at: www.iom.edu/Reports/2003/Health-Professions-Education-A-Bridge-to-Quality.aspx. Accessed April 1, 2012.
31. Jefferies A, Simmons B, Tabak D, McIlroy JH, Lee KS, Roukema H, Skidmore M. Using an objective structured clinical examination (OSCE) to assess multiple physician competencies in postgraduate training. Med Teach. 2007;29:183–91.

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9

32. Kachur E. OSCEs in the US. In: Otaki J, editor. Theory and Practice of the OSCE. Tokyo: Shinohara Shuppan Shinsha; 2007. p. 204–17. [Japanese].
33. Kachur EK, Green S, Dennis C. Written comments on Objective Structured Clinical Exam (OSCE) rating forms: an exploratory study. Teach Learn Med. 1990;2:225–31.
34. Kalet A, Earp JA, Kowlowitz V. How well do faculty evaluate the interviewing skills of medical students? J Gen Intern Med. 1992;7(5):499–505.
35. Kilminster S, Roberts T. Standard setting for OSCEs: trial of borderline approach. Adv Health Sci Educ. 2004;9:201–9.
36. Kogan J, Holmboe E, Hauer K. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. JAMA. 2009;302:1316–26.
37. Krumer A, Muijijens A, Jansen K, Dusman H, Tan L, van der Vleuten C. Comparison of a rational and an empirical standard setting procedure for an OSCE. Med Educ. 2003;37:132–9.
38. Lamont CT, Hennen BK. The use of simulated patients in a certification examination in family medicine. J Med Educ. 1972;47:789–95.
39. Morrison LJ, Barrows HS. Educational impact of the Macy Consortia: regional development of clinical practice examinations. Final report of the EMPAC Project; 1998.
40. Newble DI, Hoare J, Sheldrake PF. The selection and training of examiners for clinical examinations. Med Educ. 1980;14:345–9.
41. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrott V, Roberts T. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33:206–14.
42. Ozuah P, Reznik M. Residents' asthma communication skills in announced versus unannounced standardized patient exercises. Ambul Pediatr. 2007;7:445–8.
43. Ozuah P, Reznik M. Using unannounced standardized patients to assess residents' competency in asthma severity classification. Ambul Pediatr. 2008a;8:139–42.
44. Ozuah P, Reznik M. Using unannounced standardized patients to assess residents' professionalism. Med Educ. 2008b;42:532–43.
45. Parish SJ, Ramaswamy M, Stein MR, Kachur EK, Arnsten JH. Teaching about substance abuse with Objective Structured Clinical Exams. J Gen Intern Med. 2006;21:453–9.
46. Peabody J, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. JAMA. 2000;283:1715–22.
47. Peabody J, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. Med Care. 2004;42:1066–72.
48. Petrusa ER, Guckian JC, Perkowski LC. A multiple station objective clinical evaluation. Res Med Educ. 1984;23:211–6.
49. Pinsky L, Wipf J. A picture is worth a thousand words: practical use of videotape in teaching. J Gen Intern Med. 2000;15:805–10.
50. Poenaru D, Morales D, Richards A, O'Connor HM. Running an objective structured clinical examination on a shoestring budget. Am J Surg. 1997;173:538–41.
51. Quirk M. Intuition and Metacognition in Medical Education: keys to Developing Expertise. New York: Springer; 2006.
52. Rethans J, Gorter S, Bokken L, Morrison L. Unannounced standardized patients in real practice: a systematic literature review. Med Educ. 2007;41:537–49.
53. Reznick RK, Smee S, Baumber JS, Cohen R, Rothman A, Blackmore D, Berard M. Guidelines for estimating the real cost of an objective structured clinical examination. Acad Med. 1993;68:513–7.
54. Sadler DR. Interpretations of criteria-based assessment and grading in higher education. Ass & Eval High Educ. 2005;30:175–94.
55. Sayer M, Chaput de Saintonge M, Evans D, Wood D. Support for students with academic difficulties. Med Educ. 2002;36:643–50.
56. Schwartz RW, Witzke DB, Donnelly MB, Stratton T, Blue AV, Sloan DA. Assessing residents' clinical performance: cumulative results of a four-year study with the Objective Structured Clinical Examination. Surgery. 1998;124:307–12.
57. Segal SS, Giordani B, Gillum LH, Johnson N. The Academic Support Program at the University of Michigan School of Medicine. Acad Med. 1999;74:383–5.
58. Shah R, Edgar D, Evans B. Measuring clinical practice. Ophthalmic Physiol Opt. 2007;27:113–25.
59. Silverman J, Kurtz S, Draper JH. Skills for Communicating with Patients. 2nd ed. Oxford, UK: Radcliffe Publishing; 2005.
60. Stillman PL, Brown D, Redfield D, Sabers D. Construct validation of the Arizona Clinical Interviewing Rating Scale. Educ Psychol Meas. 1977;37:1031–78.
61. Sutnick AI, Stillman PL, Norcini JJ, Friedman M, Regan MB, Williams RG, Kachur EK, Haggerty MA, Wilson MP. ECFMG assessment of clinical competence of graduates of foreign medical schools. JAMA. 1993;270:1041–5.
62. Tamblyn RM, Klass DJ, Schnabl GK, Kopelow ML. The accuracy of standardized patient presentation. Med Educ. 1991;25:100–9.
63. Wallace P. Coaching Standardized Patients for Use in the Assessment of Clinical Competence. New York: Springer; 2007.
64. Wayne State School of Medicine Standardized Patient Program. Case Development Blueprint. Wayne State School of Medicine; 2011.
65. Yedidia M, Gillespie C, Kachur E, Schwartz M, Ockene J, Chepaitis A, Snyder C, Lazare A, Lipkin M Jr. Effect of communications training on medical student performance. JAMA. 2003;290:1157–65.
66. Zabar S, Ark T, Gillespie C, Hsieh A, Kalet A, Kachur E, Manko J, Regan L. Can unannounced standardized patients assess professionalism and communication skills in the emergency department? Acad Emerg Med. 2009;16:915–8.
67. Zabar S, Hanley K, Kachur E, Stevens D, Schwartz MD, Pearlman E, Adams J, Felix K, Lipkin M Jr, Kalet A. "Oh! She doesn't speak English!" Assessing resident competence in managing linguistic and cultural barriers. J Gen Intern Med. 2006;21:510–3.
68. Zabar S, Hanley K, Stevens D, Kalet A, Schwartz M, Pearlman E, Brenner J, Kachur EK, Lipkin M. Measuring the competence of residents as teachers. J Gen Intern Med. 2004;19(5 Pt 2):530–3.

Websites (Accessed 1 Apr 2012)

69. American Association of Colleges of Osteopathic Medicine (AECOM). www.aacom.org.
70. Association of American Medical Colleges (AAMC). www.aamc.org.
71. Association of Standardized Patient Educators (ASPE). www.aspeducators.org.
72. Centers for Disease Control (CDC). Epi Info 7. wwwn.cdc.gov/epiinfo.
73. The College of Family Physicians of Canada. Family Medicine Examination. www.cfpc.ca/FMExam/.
74. Federation of State Medical Boards (FSMB) and the National Board of Medical Examiners (NBME). United States Medical Licensing Examination – Step 2 CS (Clinical Skills). www.usmle.org/step-2-cs/.
75. Foundation for Advancement of International Medical Education and Research (FAIMER). www.faimer.org.

76. Medical Council of Canada (MCC). Qualifying Examination Part II. www.mcc.ca/en/exams/qe2/.
77. MedEdPORTAL. www.mededportal.org.
78. MedInfo Consulting. OSCE Home. www.oscehome.com.
79. National Board of Medical Examiners (NBME). www.nbme.org.
80. National Board of Osteopathic Medical Examiners (NBOME). COMPLEX L2-PE FAQ and Information. www.nbome.org/complex-pe.asp?m=can.
81. Royal College of Physicians and Surgeons of Canada. The CanMEDS 2005 Physician Competency Framework. http://rcpsc.medical.org/canmeds/CanMEDS2005/.
82. Society for Simulation in Healthcare (SSH). www.ssih.org.
83. U.S. Department of Education. Family Educational Rights and Privacy Act. www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html.
84. U.S. Department of Health & Human Services. Understanding Health Information Privacy. www.hhs.gov/ocr/privacy/hipaa/understanding/index.html.
85. Vroman Systems, Inc. FormSite. www.formsite.com.

Index

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*, https://doi.org/10.1007/978-3-031-88752-9