Organizing OSCEs (and Other SP Exercises) in Ten Steps

2

Elizabeth Krajic Kachur, Sondra Zabar, Kathleen Hanley, Adina Kalet, Julia Hyland Bruno, and Colleen C. Gillespie

Step 1

Identify Available Resources

Assemble a Team

Objective structured clinical examinations (OSCEs) and other SP projects can be a major undertaking, and as with most other educational projects, collaboration within and across specialties, even across 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 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). 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. Participating in such a team provides an opportunity to engage young, upcoming, enthusiastic faculty. Others may be involved only in the implementation phase of the OSCE (e.g., raters). Regularly scheduled meetings can help the committee become more established. After the actual OSCE, the group can work on data interpretation and dissemination.

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 education (medical school, postgraduate education, continuing medical education). Much of what is required for a successful OSCE is independent of specialty or profession.

Identify Location

When planning where to hold an OSCE, one first needs to decide how important the authenticity of the clinical environment is for the educational exercise at hand. Clinic rooms are of course the ideal spaces for OSCE stations, and one can consider scheduling the OSCE during the weekend or other time when the clinic is closed. OSCE organizers will need to work with clinical administrators well in advance, and also

E.K. Kachur, Ph.D.
Medical Education Development,
National and International Consulting,
New York, NY, USA

S. Zabar, M.D. (✉) • K. Hanley, M.D. • A. Kalet, M.D., M.P.H.
Department of Medicine, Division of General Internal Medicine,
Section of Primary Care, New York University School of Medicine,
550 First Avenue, BCD D401, New York, NY 10016, USA
e-mail: sondra.zabar@nyumc.org

J.H. Bruno
Program for Medical Education Innovations and Research,
New York University School of Medicine, New York, NY, USA

C.C. Gillespie, Ph.D.
Department of Medicine, Division of General Internal Medicine,
New York University School of Medicine,
New York, NY, USA

S. Zabar et al. (eds.), *Objective Structured Clinical Examinations*,
DOI 10.1007/978-1-4614-3749-9_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 medical school curriculum decision-making ■ Able to communicate well and create a team spirit	One 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, 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 experience ■ Is familiar with performance standards ■ Accepts editing	One or more (depending on scope)
Trainer	■ Understands SP and rater 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	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 medical case (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., conversations with faculty on break)	At least one (may not be needed if institution has a dedicated clinical skills center)
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)
Data Manager	■ Can enter performance data ■ 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/posttesting) ■ Can develop and analyze program evaluations (e.g., surveys, focus groups)	At least one

take into account technical details such as transport and set-up of props and station materials (e.g., hospital gowns, rating forms, video cameras). Some institutions are fortunate enough to have simulation centers or other dedicated training space with mock examination rooms and built-in monitoring and recording capacities. The OSCE organizer should keep in mind, however, that verisimilitude is not always necessary and learning can also be done in any classroom.

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. There may be funds available that can be used to support OSCEs. In addition, some SP programs have been funded by local medical societies, foundations (e.g., through grants for improving doctor–patient communication), and philanthropic donations.

Best Practices: Assembling a Team

- Establish a clear common goal.
- Build a team with a variety of skills.
- Schedule regular meetings to build group identity.
- Create a common repository (i.e., shared drive, secure Website) for meeting minutes, materials, and protocols.
- Look broadly for suitable sites and funding sources.

Step 2

Agree on Goals 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 Fig. 2.1 (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.

Figure 2.2 provides a list of core OSCE budget items, filled in for the same example General Internal Medicine Residency OSCE introduced in Fig. 2.1. 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 (Poenaru et al. 1997; Reznick et al. 1993).

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 amongst 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.

Figure 2.3 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 much expertise in this area, shorter planning times may be possible.

Best Practices: OSCE Planning

- Identify date and time of OSCE.
- Make a timeline working backward from the OSCE date.
- Start early to identify potential SPs and secure training times.
- Identify potential location of OSCE early (clinic rooms, class rooms, or simulation center).
- Secure participants' availability.

OSCE Project Name	Annual General Internal Medicine Residency OSCE
OSCE Goals	Assessment of general clinical competencies
Number and Type of Trainees	24 PGY-1s
Number and Type of Stations	10 independent stations with SP encounters
Potential Timing	All residents on one Saturday (two 3.5-hour sessions)
Potential Space	Outpatient department when no patient care sessions
Budget Available and Potential Funding Sources	$4,000 + $1,000 possible institutional grant
Motivational Strategies	Snacks for residents and SPs. Two rest stations per resident rotation. Immediate post-station feedback with faculty observers.

Fig. 2.1 Example worksheet for making initial OSCE plans

Budget Items to Consider	$ Needed	In Kind	Cost per Learner
Space 1 room per station, SP/faculty and learner meeting areas		Donated	
SPs – training & OSCE performance Check for local rates, costs vary depending on location and simulation task. Ask SPs to arrive ½ hour prior to the start of the OSCE and factor early arrival into payment.	4 hrs training + 2 4-hr OSCEs @ $25/hr x 10 cases (+ 2 trained alternates) $3,200		$133.33
Raters – training & rater tasks I.e., when faculty raters are used instead of or in addition to SPs; consider credit for "teaching" if direct reimbursement of faculty is not possible		Donated	
Refreshments Need not be fancy but can help create a more relaxed atmosphere	$120		$5
Medical Supply Need not be sterile but should be authentic		Donated	
Office Supply Photocopying forms, pens		Donated	
Video Equipment & Supply Sample learner performance for quality control and future learning activities. Cameras may be purchased (a recommended one-time investment!), rented or borrowed.	4 digital cameras @ $200/each + blank DVDs $840		$35
Data Entry & Report Card Assembly May be performed by in-house or temporary staff	40 hrs @ $15/hr $600		$25
Data Analysis Faculty/staff with statistical capabilities are vital to an OSCE organization team		Donated*	
TOTAL	$4,760		$198.33

* If faculty not available, additional cost to hire an outside master's level statistician estimated at $40/hr x 10 hrs = $400

Fig. 2.2 Example OSCE budget. This figure continues the example ten-station General Internal Medicine Residency OSCE outlined in Fig. 2.1. SP training time includes both rater and case portrayal training. Cost per learner is calculated for 24 residents

OSCE Project Name: Annual General Internal Medicine Residency OSCE	Date: March 19th

	Tasks	Individuals Involved	Deadline
	Overall Initial Planning		
3-4 months before the OSCE	▪ Decide on format (e.g., number of stations, time frame)	L, P, SD	Nov. 24th
	▪ Create a blueprint (identify competencies to be assessed)	L, P, SD	Nov. 24th
	▪ Decide on what to maintain from previous OSCEs/develop new cases	L, P	Jan. 1st
	▪ Identify appropriate OSCE location (stations and assembly rooms)	L, P	Dec. 3rd
	▪ Recruit staff (for administrative tasks, monitoring, time keeping)	P	Dec. 17th
	▪ Decide on SP/rater recruitment and training schedule	P	Dec. 17th
	▪ Communicate with learners (e.g., provide dates/times, explain procedure)	L, P	Jan. 7th
	▪ Clarify budget (e.g., SP costs, refreshments)	P	Dec. 17th
	▪ Consider videotaping	L, P	Dec. 17th
	Station/Material Preparations		
1 week - 3 months before the OSCE	▪ Review old station/OSCE materials (e.g., learner/SP instructions, rating forms)	L, P, SD, A	Jan. 14th
	▪ Develop new station/OSCE materials (i.e., content generation and formatting)	L, P, SD, A	Jan. 31st
	▪ Determine SP payment process	P, A	Dec. 17th
	▪ Make room arrangements	A	Dec. 17th
	▪ Recruit SPs	P	Feb. 18th
	▪ Train SPs	*!	Mar. 11th
	▪ Recruit faculty	L, P	Jan. 14th
	▪ Prepare faculty (e.g., circulate station/format information, rater training)	P	Feb. 18th
	▪ Prepare props (e.g., fake pill bottles and charts)	P	Mar. 4th

Fig. 2.3 Example worksheet for assigning OSCE responsibilities and creating timelines. "Individuals Involved" follow the OSCE staffing roles listed in Table 2.1 (*L* leader, *P* planner, *A* administrator, *SD* station developer, *Tr* SP trainer, *Ti* timer, *M* monitor, *DM* data manager, *DA* data analyst)

General Preparations

1-2 weeks before the OSCE	▪ Order supplies (e.g., paper, folders)	A	Mar. 4th
	▪ Order refreshments	A	Mar. 14th
	▪ Photocopy station materials	A	Mar. 11th
	▪ Assign SPs, faculty and learners (create assignment sheets)	A	Mar. 4th
	▪ Prepare name tags/labels for learners (assign learner IDs)	A	Mar. 11th
	▪ Develop rotation schedules	A	Mar. 4th
	▪ Prepare invoices and necessary paperwork for SP payment	A	Mar. 4th
	▪ Prepare signs (e.g., station numbers, arrows to signal flow)	A	Mar. 18th
	▪ Orient hall monitors and time keepers	P	Mar. 18th

OSCE Administration

day of the OSCE	▪ Prepare stations and assembly rooms (signs, station materials, refreshments)	M	Mar. 19th
	▪ Assign substitutes (if necessary)	M	Mar. 19th
	▪ Orient faculty, SPs and other personnel	P	Mar. 19th
	▪ Position faculty, SPs, hall monitors and time keepers	M	Mar. 19th
	▪ Orient learners (e.g., assign starting station, disseminate name/number labels)	L	Mar. 19th
	▪ Guide learners to individual starting stations	M	Mar. 19th
	▪ Time stations (start, feedback, station changes)	Ti	Mar. 19th
	▪ Manage emergencies (e.g., equipment breakdown)	M	Mar. 19th
	▪ Assure smooth changeovers of SPs, faculty and learners	M	Mar. 19th
	▪ Reassemble learner group (e.g., for debriefing, program evaluations)	L	Mar. 19th
	▪ Collect and count all forms	A	Mar. 19th
	▪ Clean up stations and assembly rooms	M	Mar. 19th

Post-OSCE Tasks

days to months after the OSCE	▪ Sort out forms	DM	Mar. 21st
	▪ Ensure timely SP payment	A	Mar. 25th
	▪ Enter data and evaluation results	DM	Mar. 25th
	▪ Analyze data	DA	Apr. 8th
	▪ Report evaluation data (e.g., report cards)	P, L	Apr. 19th
	▪ Report on experience internally and externally (e.g., presentations, articles)	L, PE	Sept. 19th

Fig. 2.3 (continued)

Step 3

Establish a Blueprint

A key element for designing an OSCE is the development of a blueprint. This is a matrix that connects a list and brief description of all stations with the competencies that are being assessed (see, e.g., Fig. 2.4; a blank blueprint is included as Appendix D). This ensures that individual competencies are examined multiple times and that each station contributes to the overall comprehensiveness of the exam or exercise by assessing multiple competencies. We create our blueprints by selecting cases from our case bank (see Fig. 2.10), a useful repository which organizes our accumulating cases by key blueprinting information, tracks case usage, and enables tailored querying (e.g., distribution by age, percentage New, Ongoing, Follow-up, and Discharge cases).

An organized approach to blueprinting strengthens an OSCE's validity. This can include literature reviews, curriculum surveys, and consensus building discussions. OSCEs should provide a good cross section of medical encounters typically experienced by learners. If the OSCE is a formative exercise, post-OSCE feedback from trainees (see Appendix K for a participant post-OSCE survey) should confirm that the stations assessed issues they encounter in their current work or are likely to encounter in their future practice.

The final station sequencing 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).

Once a first draft of a blueprint is completed, organizers should ask themselves the questions listed in Table 2.2.

Table 2.2 Questions important 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 races and cultural backgrounds?

Best Practices: Blueprinting

- Delineate core competencies.
- Establish performance criteria for each level of training.
- Ensure OSCE case patient age, gender, race, and prevalence of disease reflect actual clinical practice.
- Align OSCE skills and content assessed with current or new curricula.

	Station		Skills Assessed				Comments
	Case	Content Areas	Comm	History Gathering	Physical Exam	Treatment Mgmt & Plan	
1	Urethritis Follow-up - take sexual Hx from bisexual pt	Prevention (HIV, sexual Hx, STI prevention)	X	X		X	M. Unlink from lab tests.
2	High Risk Smoker - move pt from contemplation to action stage, develop plan	Addiction Medicine (smoking, behavior change cycle)	X	X		X	M. Adjust case for local pt population.
3	Street Fair - counsel pt about positive bone density results	Geriatrics (osteoporosis, bone density, Dexa interpretation)	X	X		X	F. Clarify instructions, change to Dexa results.
4	Diabetes (phone) - triage problem over the phone	Common Symptoms/ Undifferentiated Problems (DDx of diarrhea), Phone Medicine	X	X		X	F. Adapt case from other program.
5	Difficulty Sleeping - screen appropriately	Common Symptoms/ Undifferentiated Problems (PTSD, DDx of sleep problems)	X	X		X	M/F. Develop new case.
6	Loss of Loved One - counsel pt who recently lost spouse	Common Symptoms/ Undifferentiated Problems (stages of grief, grief counseling)	X			X	F. Emphasize common symptoms or acute problems?
7	Diabetes Precepting - precept SL, new onset diabetes case	Acute Problems (diabetes), Microskills Teaching		X		X	M/F.
8	Asthma - assess tightness in chest, counsel re: medication	Acute Problems (asthma Dx and Tx)	X	X	X	X	F. How many residents got prednisone training?
9	Test Results (phone) - liver abnormalities	Acute Problems, Lab Tests, Phone Medicine	X	X		X	M. Check new recommendations.
10	Teaching Px Skills - teach student Px procedures	Px Procedures, Bedside Teaching (incorporating pt)	X		X		Similar to last year except for procedure.

Fig. 2.4 Example blueprint for an Internal Medicine residency OSCE (*Hx* history, *Px* physical exam, *Dx* diagnosis, *DDx* differential diagnosis, *Tx* treatment, *STI* sexually transmitted infection, *PTSD* posttraumatic stress disorder, *SL* standardized learner)

Step 4

Develop Cases and Stations

A blueprint leads to a profile for each of the stations which then can serve as a starting point for case development (the *case* is the clinical problem; the *station* involves a specific set of tasks being assessed in the OSCE). Basing OSCE stations on real patient cases will add validity. However, after disguising the identity of the source patient, it may be necessary to make adjustments for the training level, OSCE focus, or the time limitations imposed by the exercise. Figure 2.5 illustrates how one can adjust 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.

Educators should not feel obligated to start from scratch in developing their OSCE cases. Our case development worksheet is included in Appendix I. (Our template follows Silverman et al.'s (2005) History of Present Illness framework and was refined with reference to the Wayne State School of Medicine Standardize Patient Program's (2011) case development tool.) See also Appendix J for a checklist development worksheet. Additional selected station/case development resources are included in Appendix P (Other Resources). We also recommend reaching out to other health professions schools; many programs will likely be willing to share their OSCE cases.

As part of the station development process it is important to try out new cases through role-play and adherence to the given time limits. Sometimes multiple enactments are necessary to gain clarity on issues such as scope of task or SP emotional tone. Role-play at this stage should involve faculty who know the target learner group and the sort of questions they are likely to ask the SPs.

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. Table 2.3 provides considerations specific to each component of the documentation accompanying each case. A sample case (including station overview, directions for the OSCE participant, and detailed SP case portrayal instructions) plus corresponding rating forms for both the SP and faculty observer are provided in Appendices F, G, and H.

	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

Communication Barriers

- language
- literacy/health literacy
- cultural differences (e.g., dress code)
- altered cognitive state (e.g., intoxication, dementia, confusion)
- intense emotional state (e.g., depression, anger, mania, anxiety, shock, shame)
- difficulty/inability to speak/hear/see
- physical setting (e.g., patient in bed, connected to hospital equipment)
- telephone/online communications
- multiple interviewees/interviewers

Fig. 2.5 Adjusting a case for station difficulty

Table 2.3 Overview of station-specific materials (their purpose, content, and special considerations)

Forms	Purpose	Content elements	Considerations/tips
Station overview	To assist program organizers	■ 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)
Learner instructions	To communicate the scenario and tasks to learners before they enter the station	■ Patient information (e.g., name, age, occupation) ■ Reason for visit ■ Learner role ■ Starting point for encounter (beginning, middle, end) ■ Situation (medical/psychosocial information available, prior developments/encounters) ■ Learner task(s)	■ Be brief (consider reading time) ■ Assure equal length with other stations ■ Timeline with arrows can help orient learner quickly ■ Bulleted information can be read faster ■ Use language learners are familiar with (e.g., well-known abbreviations)
Fact sheets (only 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	■ Guidelines for diagnosis and treatment (case specific) ■ 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) ■ Assure parity with other stations ■ Organize material to be reviewed quickly ■ Use graphs where possible ■ Assure accuracy ■ Avoid controversy
SP instructions	To prepare SPs for their case	■ Scenario (what happened from the SPs 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 for learner (verbal, nonverbal) ■ Timing (beginning, middle, end/after 2-min warning)	■ Provide opportunity for SPs to personalize scenario within limits (e.g., name of spouse) ■ Supply an "opening line" and specific messages to give ■ Be specific ■ 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 challenges for the learner/station goals
Rating form	To capture the performance assessments	■ Administrative information (e.g., learner IDs, date, station) ■ Dimensions on which to assess the learner (e.g., communication skills, case management) ■ Checklist or global rating items ■ Room for comments (e.g., areas of strengths, areas in need for improvement)	■ Make items evidence based ■ 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 ■ Pretest for readability and ability to observe and rate ■ Include at least one summary rating for cross-validation
Faculty instructions	To standardize faculty assessment and teaching	■ Procedural steps for observing encounters (e.g., positioning to observe nonverbal behavior, start/stop video) ■ 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 ■ Assure that teaching points match the station goals
Post-encounter materials (optional)	To give learners the opportunity to reflect on/synthesize the encounter, receive feedback, or extend their clinical reasoning about the case	■ Patient note (with space for summarizing history, diagnosis, and treatment plan) ■ Supplementary diagnostic test results (e.g., EKG, X-ray)	■ Be selective and pragmatic: e.g., weigh faculty availability for giving feedback versus gathering further learner data ■ Consider computer- versus paper-based ■ Consider reserving the time between stations for rest with no post-encounter activities

Table 2.4 Guidelines 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 appropriate)*
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?"

Table 2.5 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 and what learners need to be able to do?
■ *Is the case appropriate for the learner?* Consider profession, training level, course/rotation content
■ *Can the tasks be managed or at least initiated in the given time?* (e.g., 10 min)
■ *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 is supposed to 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?* For example, can someone stay who depressed for 10 min? Will the case require multiple SPs because it is too stressful or too difficult to maintain a particular physical finding?
■ *Will it be possible to find an adequate number of SPs to portray this case?* If not, can age, gender, or other characteristics be changed to make the search easier?

Each OSCE form should be clearly marked with station number and title. The title needs to be phrased in a way not to give away the sometimes hidden, station-specific challenge (e.g., "Secret Drinker"). 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, receiving immediate feedback, or, simply, rest. Which option one selects will depend on one's goal for the OSCE and the station, as well as pragmatic considerations such as faculty availability to observe and debrief encounters (see Table 2.3). 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 2.4 provides a sample set of instructions 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 are accomplished with each learner. Also see the feedback training protocol (Table 2.11).

To assure the quality of each case, organizers should ask themselves the questions listed in Table 2.5.

Best Practices: Case Development

- Choose scenarios that are both common and challenging presentations for your learners.
- Ensure that cases represent the patient population in your clinical environment.
- Build specific goals and challenges into each scenario.
- Choose a post-encounter activity (i.e., feedback, supplementary exercise, or rest).
- Make sure it is possible to complete tasks in the time allotted.
- Organize a trial run with a variety of other learners to validate and fine tune cases.

Step 5

Create Rating Forms

The quality of a rating form is judged by the degree to which raters, both SPs and/or faculty, can use the form consistently (i.e., reliability, the degree to which the form would produce the same results if used by different raters or on different occasions) and the degree to which the elements of the rating form accurately reflect the intended skills and performance (i.e., validity). The keys to developing reliable and valid rating form items are (a) identifying the specific domains, (b) writing understandable items, and (c) providing anchors or instructions that guide raters in their assessment. By establishing a blueprint which specifies what skills and content the OSCE is designed to assess, and how each station contributes towards this goal, much of the work in creating effective rating forms is already done. The items in the rating form should reflect the blueprint and can therefore include both skills assessed across all stations within an OSCE as well as content and skills specific to a station or subset of

stations. Two formats for the rating form items are typically used: behavior-specific items (did the trainee perform a specific behavior?) and global ratings. Both are important (Norcini et al. 2011) and many rating forms include both. These two types of items are usually strongly correlated; however, each may provide unique information about trainee performance. From the perspective of feedback on performance, behavior-specific checklist items provide learners with actionable data while global ratings are much less directive. Space for comments is also useful to provide opportunities to indicate rating challenges or more specifics about the learner's performance (Kachur et al. 1990). Sample rating forms can be found in Appendices G, H, and O. Appendix J provides a checklist development template.

Behavior-Specific Rating Form Items

Table 2.6 provides a stepwise process for developing behavior-specific rating form items. The number of behaviorally anchored items that are assessed within a particular domain affects the quality of the measurement. The more items, the more reliable and valid a rating form is likely to be. The trade-off is the burden on the rater. Asking raters to rate too many and/or very complex aspects of performance can lead to decreased accuracy. Extensive, targeted training of raters and providing adequate time to rate are two additional ways of achieving a good balance.

Checklists are popular in OSCEs because of their simplicity—noting simply whether or not specific behaviors or actions were performed can enhance the accuracy and reliability of ratings. However, such simplification may miss important dimensions of performance and could, in some circumstances, compromise the validity of the assessment tool. 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. Thus scales that provide multiple rating options (e.g., Likert-type or 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, may take up valuable rating time, and can lead to a reduction in reliability if the response options don't align well with learner behavior.

One compromise is to use a trichotomous anchoring system, such as "not done," "partly done," "well done." This approach seems to help overcome the tendency of many raters to "give credit" to learners whenever possible and also helps set a high standard for performance. Looking ahead at the interpretation of performance data which will result from the OSCE, one can then 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 in order to avoid floor (everyone does poorly) and ceiling (everyone does well) effects and maximize the degree to which the items differentiate among learners.

Table 2.6 Stepwise 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
2.	*Compare that conceptualization with available standards* e.g., 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., done/not done checklist versus scale
5.	*Create behavioral anchors to help evaluators identify which rating option to select* e.g., 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 physician to a family member or friend, reflecting measures widely used with "real" patients to assess patient satisfaction and quality of care.

Global ratings are often thought to be less reliable because they are not anchored in specific, observable behavior and therefore more susceptible to the subjectivity of the rater. However, in certain situations—with "expert" raters evaluating more holistic competence—global ratings may be more valid than checklist ratings.

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 the behavioral-specific items. Generally the reliability of global ratings has been quite satisfactory (Hodges & McIlroy 2003). Often such ratings use a 4- or 5-point scale and have specific anchors. For example, recommendation ratings ("would you recommend this physician to a family member or friend?") can use simple descriptions such as "not recommend," "recommend with reservations," "recommend," "highly recommend," or more specific, complex descriptions of exemplar levels of performance that constitute each point on the scale. Global ratings require less attention to performance details and thus less memorization for SPs. On the other hand, there are many (often uncontrolled) factors that influence them, including subjective

biases like halo, availability, social comparison, and selective attention biases. In addition, formative feedback is difficult to provide on the basis of such global ratings.

Best Practices: OSCE Checklists

- 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 achieve a balance in terms of helping raters reflect important elements of their subjective responses and to enhance their objectivity in representing what happened during the encounter and providing learners with specific and more holistic feedback.
- Develop response options for behavior-specific items that reflect observable actions and strive to match the response options to the likely variation in performance of the learner population to maximize differentiation.

Step 6

Recruit and Train SPs

Recruitment

Think of choosing SPs as a theater director would cast a show. Each case has unique requirements, some are physiological, others are psychological. Before starting with the recruitment process it is helpful 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 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, inability to express emotions if pain is of importance in the case, or a hostile interpersonal approach if the case asks for a withholding attitude. Casting the right person for the case is important for creating an appropriate degree of realism. Even when they are experienced actors, it is difficult for SPs to overcome their typical ways of behaving or expressing themselves. If a person is exceptionally outgoing and actively expressing emotions through nonverbal behavior, then a case where tiredness and lethargy are the issue may be less appropriate. The energy to transfer a very active style into a passive one may distract from other tasks such as remembering the history items or evaluating the trainee.

Table 2.7 SP characteristics that simplify training

SP characteristic		Effect on training
Acting experience	➪	Less need to train acting (especially of high emotional levels)
Health care professionals (or trainees in the health professions)	➪	More understanding of learner role and technical issues (e.g., interview, physical exam)
No personal expertise with the case problem	➪	Less emotional involvement with the case
Personal experience with the case problem	➪	Disease-related knowledge is already present
Type casted	➪	Less need to teach affect
Prior SP experience	➪	Less need to teach the mechanics of OSCEs
Use of SPs own background	➪	Less history information to remember
Over age 18	➪	No need for developmental considerations
Under age 70	➪	Easier to train, may remember better
GTA or UTA experience	➪	Comfortable with physical exams, used to focus on performance details, expert in breast and pelvic or urological exams

GTA gynecological teaching associate, *UTA* urology teaching associate

Familiarity with the medical problem in focus can either help or hinder the simulation. On one hand, having experienced a medical problem oneself may provide special insights into the case. On the other hand, memories about 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. To avoid an increased need for SP maintenance, it is better to select SPs for whom the medical problem in focus does not evoke special memories. As Table 2.7 illustrates, by looking ahead at training requirements one can consider some SP characteristics that are likely to reduce the need for preparations.

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 cannot explode on the examinees 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.

Actors have been viewed by many as ideal candidates. 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. Not every actor is willing to go along with that, and often times a real acting opportunity will be preferable to taking on an "educational" role. Thus the OSCE project can quickly be missing an SP.

Once a program has developed a cadre of SPs, word of mouth will often become the most effective and efficient way of recruitment. One experienced SP coordinator felt that a referral from another SP has a one in two chance of bringing in a good candidate, with a physician referral the chances are one in three. Using ads, only 1 in 20 responses may lead to hiring (Tamblyn et al. 1991).

Training for Case Portrayal

To make a patient's case come to life SPs need to become accomplished in three different areas. (1) They must know all the physical, psychological, and social details related to 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. (3) Their actions and responses must be timed correctly. Many novice SPs tend to give away all the information they have about the case 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 information prematurely reduces the learner's chances to work on important skills. 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. Thus SPs 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. They can even view a standard setting videotape to emphasize nonverbal behavior 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 2.8 lays out a protocol for training SPs. For logistical reasons or time limitations it may not always be possible to go through all steps, but one could consider those in the shaded boxes as the most essential ones. There are varied opinions as to how much training is necessary for SPs to perform their case adequately. A relevant 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; (5) Final Dress Rehearsal (Wallace 2007). The total amount of training time will depend on case requirements,

Table 2.8 Training protocol: SP portrayal

1. *Provide training program overview* e.g., when and how to get where, who will they be working with, what are the program objectives, what is the history of the project, what will a typical encounter with learners be like, who else will be in the room, what prior experience learners will have had with OSCEs/SPs
2. *Explore SP expectations/concerns* e.g., prior work with learners at the targeted training level—how did it go, how did it compare to their expectations, what are their concerns, what are they looking forward to, how might it be similar/different from previous SP work
3. *Review individual cases* break into subgroups, have SPs take turns reading aloud the learner instructions, SP instructions, and rating form, 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, home address), 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 or only upon prompting by the learner
4. *Review video sample encounter* to get at emotional tone, nonverbal behavior, bring out more of SPs past simulation experiences, show learner's expected level of performance
5. *Demonstrate SP encounters* select SP volunteer or SP who has portrayed same or similar case before, others watch while referring to SP instructions and rating forms, time the encounter as you would during the OSCE, model what would happen during and after the encounter (e.g., physical exam, rating, feedback), discuss case portrayal, recheck the SP instructions if indicated, if there are multiple demonstration interviews change SPs and modify interviewer approach (e.g., poor performance, unprofessional behavior)
6. *Videotape practice encounters and review performance* reviews can be done in a group or SPs can watch tapes independently and then discuss their impressions and reactions
7. *SPs practice with each other* make sure everyone takes on the learner role at least once to appreciate the challenges involved in the case, reduce anxiety performance by "requesting" the interviewing SP to make mistakes as a learner might, include rating and feedback to learner in role-play (if applicable)
8. *Organize trial runs* the more practice, the more SPs will learn about potential learner approaches to the case (e.g., questions, physical exam maneuvers), organizing a mock OSCE (if possible in the place where the real OSCE will occur) can provide unique practice opportunities and greatly enhance understanding of context and timing

cost, and time limitations. If it is a formative assessment 2 h may be adequate, especially with SPs who have experience. If it is a summative assessment, training will have to be much more extensive and, there are literature reports of 10–20 h of training (ibid). However, the latter will have to be divided into shorter training segments. Typically 2 h is a limit to how much SPs can absorb at one time. We typically train SPs for 4–6 h, including a minimum of 2 h focusing on the case and 2 h on the checklist. When organizing a higher-stakes event one must definitely consider a trial run. New SPs can especially benefit from getting a first-hand experience of the tasks and timing involved.

Best Practices: SP Recruitment and Training

- Search for SPs through word-of-mouth strategies (e.g., by contacting other SPs, connecting with other SP trainers, talking to clinicians and acting teachers).
- Cast the right person for each case (i.e., physical appearance, psychological profile, availability, no contraindications).
- 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).
- Put SPs into learner's positions through role-play to enhance their understanding of the case (e.g., interactive and emotional impact of SP actions) and to promote an empathic approach to learners.
- Practice all aspects of the encounter (e.g., physical exam, feedback); do not leave SP performance to chance.
- 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).
- Train all SPs who are portraying the same case (simultaneously or consecutively) at the same time to enhance consistency in case portrayal across SPs.

Step 7

Recruit and Train Evaluators

An important decision when planning an OSCE is who will rate the participants. Depending on the OSCE project, faculty, SPs, and/or peers will be entrusted with the responsibility of rating a trainee's performance. At times evaluations are completed by more than one group of observers. Often organizers do not have the luxury to select raters even though some research suggests that recruiting the right people might be more important than training them (Newble et al. 1980). An initial rater screening strategy could consist of assembling candidates in small groups and showing them selected videotapes 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. On the surface, faculty raters may appear ideal, but they are not necessarily accurate (Kalet et al. 1992) 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 when signing them on for SP work.

Regardless of whether the rating is done by SPs, faculty, or peers, attention must be given to raters providing as accurate and reliable ratings as possible. The rater task is difficult because there are so many factors that can interfere with an accurate 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 at each of these rater tasks as illustrated in the rater self-assessment guide in Table 2.9.

Raters need to be aware of their rating style, whether they are "doves" (i.e., easy raters) or "hawks" (i.e., harsh raters), and what types of errors they are more likely to make. Self-awareness is no guarantee of being completely error free, but it is the best chance to provide a fair rating.

If possible, raters should be trained groups. A rater training protocol is detailed in Table 2.10. 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 too, need some type of orientation, if necessary in writing, to orient them to the goals, process, and content of the exercise.

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 exam development, raters must at least understand the underlying rationale and feel confident that categories were not selected arbitrarily. Rater trainers have to 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, even worse, when the project is over and one realizes that a rating form item has been completely misunderstood.

Table 2.9 Helping raters improve their accuracy (rater self-assessment guide)

	Key question	WHAT I NEED TO WATCH OUT FOR:
Observation	*What knowledge, skills and attitudes did I observe?*	■ *Too little, too much, or selective attention to details* inappropriate focus ■ *Halo effect* one observation which is easy to obtain or of great significance to rater influences perception of other behavior—first impression error ■ *Observation is too short or too long* premature closure or loss of information
↓		
Judgment	*How should I rate this trainee on this item?*	■ *Gravitation towards the mean or extremes* central tendency/end-aversion bias or overused end scale points result in too little or too much range ■ *Similar-to-me effect* trainees more similar to rater receive better scores ■ *Contrast effect error* trainees are evaluated against each other and not against an external standard ■ *Generalizations, prejudices, and stereotyping* ■ *Standards are not fully understood* unclear about expectations for training levels ■ *Differences between rating scale points are unclear* ■ *Rater style*: __ dove, __ moderate, __ hawk ■ *Mum effect* hesitation to provide poor performance ratings
↓		
Documentation	*How do I complete the rating form?*	■ *Incorrect recording* evaluation judgment is not properly marked off ■ *Inadequate or missing comments*

Table 2.10 Training protocol: rating

1. *Provide training program overview* e.g., when to get where, who will they be working with, what are the program objectives, what is the history of the project, what will a typical encounter with learners be like, who else will be in the room, what prior experience learners will have had with OSCEs/SPs
2. *Explore rater expectations/concerns* e.g., prior work with learners at the targeted training level—how did it go, how did it compare to their expectations, what are their concerns, what are they looking forward to, how might it be similar/different from previous rater work
3. *Review case to be observed and rated*
   - Provide a copy of the rating form and define each item (providing examples for the response options)
   - Provide all other case materials (including learner and SP instructions)
   - Let the rater take on the role of a learner to get a personal experience of the case challenges
4. *Perform practice ratings*
   - Use live encounters or videos to demonstrate a "gold standard" evaluation 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)
5. *Review typical rater errors* discuss factors that can interfere with rating tasks (see self-assessment form in Table 2.8), encourage raters to become aware of their own style and tendencies
6. *Introduce raters and SPs* (if rating is done by a faculty observer)
   - Encourage them to work together without sharing their individual impressions about the learner's performance before documenting their own ratings
   - Give raters and SPs time to be alone to get to know each other before the first learner arrives
   - Request that they play through the case with the rater taking on the learner role to build understanding and empathy

Table 2.11 Training protocol: feedback

1. *Provide a feedback framework* • 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* • Learner self-assessment first • 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 • Connect station 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
4. *Practice giving feedback* e.g., utilizing a video of a performance and role-play

Often in many OSCEs, raters are also asked to provide immediate feedback. Typically there are time limitations (5–10 min) and feedback providers need to be brief. Table 2.11 provides a sample protocol 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 are accomplished with each learner. Providing succinct and meaningful feedback is not always that easy. If raters are also expected to give feedback they should practice doing so in advance of the OSCE (Hatchett et al. 2004).

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

- Select evaluators who are willing to adopt the program values, who are consistent in their ratings and don't have an ax to grind.
- 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).
- Make raters aware of potential biases and rating mistakes.
- Provide written guidelines for rating items, evaluation scheme, and station objectives/teaching points.
- Post-OSCE, give feedback to raters about how their ratings compare with those of others (e.g., more or less lenient, lack of range).

Step 8

Implement the OSCE: Managing the Session

In addition to station-specific materials, it is also necessary to develop forms and other resources that help with the overall organization of the event. Table 2.12 details the various forms that will be needed. Figure 2.6 provides an example station rotation schedule for OSCE participants, and Fig. 2.7 shows the same schedule from the perspective of the SP/rater. The OSCE participant schedule is also included in blank worksheet form in the back of this book (Appendix E) along with program evaluation surveys (Appendices K–M).

Whenever one plans an event that involves a large number of people, organization can be challenging. One must accept the fact that irregularities will occur, but with good planning and adequate resources, one should be able to make the program manageable. To make trouble shooting at the time of the OSCE easier, it is helpful to contemplate potential solutions ahead of the event. Key concerns include attendance, standardization, time and emotion management. Organizers should ask themselves what they could do in the event of the contingencies listed in Table 2.13. We have included some solutions that have worked for us. By having extra SPs and faculty on hand one can overcome lateness or absences. Adequate training, extra props, and forms can help with standardization. Small time and personnel adjustments may be necessary to keep the OSCE on schedule. Organizers and monitors need to be on the lookout for nervous learners who

Table 2.12 General OSCE materials

Forms	Purpose	Content elements	Considerations/tips
Learner orientation materials	To record attendance and assign ID codes (if applicable)	■ OSCE name, location, date ■ Learner names and ID codes	■ Provide consent forms (if appropriate)
SP/rater orientation materials	To record attendance and match SP/rater names with ID codes (if applicable)	■ Location, date, OSCE number ■ SP/rater names and ID codes	■ Permit room for multiple SPs per station if alternates ■ Allow room for comments and to record special occurrences ■ Provide forms for SPs or others to receive payment
Rotation schedules	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, location, date ■ List of participant names/IDs ■ Areas for indicating rotation 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 special occurrences ■ Provide room for monitor(s) name(s)
Learner post-OSCE program evaluation forms	To evaluate the OSCE	■ Self-assessment of performance ■ Prior exposure to clinical tasks/cases ■ Emotional reaction to stations ■ Realism of stations ■ Representativeness of performance ■ Motivation to perform well	■ Keep it brief ■ Comments can provide interesting qualitative data
SP and faculty program evaluation forms	To evaluate the OSCE	■ Level of case difficulty ■ Educational value ■ Faculty development value (if faculty rating) ■ SP performance (if faculty rating) ■ Level of enjoyment ■ Appropriateness of case ■ Effectiveness of instructions	■ Keep it brief ■ Comments can provide interesting qualitative data

Table 2.13 OSCE troubleshooting: potential problems and possible remedies

What if…
■ *Someone doesn't show?* For high-stakes OSCEs, always cast and train extra SPs. Consider scheduling open slots into the participant exam schedule to accommodate unforeseen emergencies. For formative OSCEs, ask a faculty member to portray the patient
■ *Someone has to leave temporarily?* Participants and SPs should be informed in advance when designated breaks will occur. For long exams 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?* Designate a staff member to regularly review and count all forms during the OSCE so rating errors can be corrected in real time
■ *A participant enters the wrong station?* Make sure exam proctors are monitoring the exam and can make timely substitutions in the trainee rotation schedule
■ *Timing is off-schedule?* If a station goes overtime, try shortening subsequent rotations by small increments until the schedule is back on track
■ *Someone is late or has to leave early?* Again, make sure time expectations are clear, and prepare back-up SPs
■ *An SP does not portray the case correctly?* Schedule ample training so that everyone is happy with the case portrayal before the actual OSCE. Make sure there is a staff member familiar with the cases present at the OSCE to answer any questions of SPs that may arise in student encounters. Consider videotaping OSCE stations for quality-control post-OSCE
■ *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
■ *Some stations consistently take less than the allotted time?* Check in with the SP between rotations; adjust details of the case portrayal if needed (This is not necessarily a problem)
■ *The OSCE is running out of time?* A participant's "score" in an OSCE is based on his or her performance in multiple stations and should not be compromised as a result of exam scheduling delays. Try first to see if SP and participant can stay late to finish the OSCE

Rotation Schedule for: General Internal Medicine Residency OSCE
Date: March 19th (PM Session)

	Rotation (# & start time)											
Station #	1	2	3	4	5	6	7	8	9	10	11	12
Participant ↘	1:00	1:18	1:36	1:54	2:12	2:30	2:48	3:06	3:24	3:42	4:00	4:18 *
Dr. A	1	2	3	4	5	rest	6	7	8	9	10	rest
Dr. B	2	3	4	5	rest	6	7	8	9	10	rest	1
Dr. C	3	4	5	rest	6	7	8	9	10	rest	1	2
Dr. D	4	5	rest	6	7	8	9	10	rest	1	2	3
Dr. E	5	rest	6	7	8	9	10	rest	1	2	3	4
Dr. F	rest	6	7	8	9	10	rest	1	2	3	4	5
Dr. G	6	7	8	9	10	rest	1	2	3	4	5	rest
Dr. H	7	8	9	10	rest	1	2	3	4	5	rest	6
Dr. I	8	9	10	rest	1	2	3	4	5	rest	6	7
Dr. J	9	10	rest	1	2	3	4	5	rest	6	7	8
Dr. K	10	rest	1	2	3	4	5	rest	6	7	8	9
Dr. L	rest	1	2	3	4	5	rest	6	7	8	9	10

*Last station ends at 4:36 pm; debriefing with faculty and residents after the last station

Fig. 2.6 Example OSCE participant rotation schedule. Shown here are the order of rotations (including two rest periods) for half of the 24 residents in our example ten-station OSCE. Each 18-min rotation period includes 5 min feedback. Participant ID numbers may be substituted for names where confidentiality is required (e.g., in a higher-stakes OSCE)

Rotation Schedule for: General Internal Medicine Residency OSCE
Date: March 19th (PM Session)

	Rotation (# & start time)											
Participant	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Station ↘	1:00	1:18	1:36	1:54	2:12	2:30	2:48	3:06	3:24	3:42	4:00	4:18
1 – Urethritis follow-up	A	L	K	J	I	H	G	F	E	D	C	B
2 - Smoker	B	A	L	K	J	I	H	G	F	E	D	C
3 – Street fair	C	B	A	L	K	J	I	H	G	F	E	D
4 – Diarrhea (phone)	D	C	B	A	L	K	J	I	H	G	F	E
5 – Difficulty sleeping	E	D	C	B	A	L	K	J	I	H	G	F
6 – Loss of loved one	G	F	E	D	C	B	A	L	K	J	I	H
7 – Diabetes precepting	H	G	F	E	D	C	B	A	L	K	J	I
8 - Asthma	I	H	G	F	E	D	C	B	A	L	K	J
9 – Test results (phone)	J	I	H	G	F	E	D	C	B	A	L	K
10 – Teaching Px skills	K	J	I	H	G	F	E	D	C	B	A	L

Fig. 2.7 Example OSCE SP/rater rotation schedule. This figure presents the same schedule as in Fig. 2.6, now highlighting the order of residents passing through each station

Best Practices: Optimizing the Test Environment

- Conduct a "dress rehearsal" prior to any high-stakes OSCE.
- Come prepared with extra forms and knowledge of office facilities (computers, printing, copying) near the testing site.
- In designing the OSCE rotation schedule, include time for orientating learners and SPs, as well as time between scenarios and after for post-OSCE debriefing.
- SPs can optimally perform and rate for up to 180 min. There should be a break if you are doing two OSCE sessions on 1 day.
- If the location is not a simulation center then testing staff should include one proctor for each clinical area (e.g., hallway) and one overall administrator.

may enter stations too early or tired SPs who do not portray their case correctly anymore. Post-OSCE debriefing will be useful for all involved.

Step 9

Manage, Analyze, and Report Data

It's 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 uninterpretable 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 planning accordingly. We have found that good data management practice—which includes protecting trainee privacy—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 early in the planning process.

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. Table 2.14 provides 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 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 memory of the logistics are still fresh (e.g., data are missing because someone arrived late).

While data can be initially entered into a spreadsheet (e.g., Microsoft Excel), which is familiar to most people, we recommend the use of data entry forms that facilitate fast, consistent, and error-free data recording that are easily exported into 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., Epi Info [wwwn.cdc.gov/epiinfo]; FormSite [www.formsite.com]).

Field-based data entry also facilitates the creation of a "data dictionary" that provides information on each data item (e.g., the checklist item it represents and in which case it was asked), how the response options were entered (e.g., 0 = no; 1 = yes; or 1–4 for global ratings), the identity of the raters (often good practice to develop an ID system for identifying the SPs), and any issues or problems that should be noted relevant to the OSCE. It's 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.

Analyzing OSCE Data

Start with descriptive statistics such as distributions of ratings across the response categories (frequencies) for each item on

Table 2.14 Creating a secure and confidential OSCE data storage system

1.	*Generate a unique ID for each individual learner* e.g., 4-digit number
2.	*In a two-column table, link these new IDs to learners' names and other identifying information* (e.g., email address, schools attended, system IDs)
3.	*Store hard and electronic copies of the table in secure locations* e.g., password-protected database file, locked file cabinet; limit access to those with responsibility for learner assessment
4.	*Store OSCE data with the unique ID ONLY* i.e., delete all other identifiers
5.	*Create a regular system for backing up your data*

the checklist to identify data entry errors and missing data. Then, once you feel the database is accurate, summarize the data across learners to identify program-level gaps in training for specific skills and to establish norms for the group (see Fig. 2.8). 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 the checklists.

Calculating and Interpreting OSCE Scores

The reasons for calculating OSCE scores are: (1) To set minimum standards for high stakes, pass/fail examinations; (2) To provide feedback to learners (and their faculty) on performance; and (3) To provide overall feedback to your program on the effectiveness of training.

Scores can be based on averages of scaled items or on percentages; the latter are used especially for checklist scores (e.g., percentage of behaviors "done"). If scaled items are non-normally distributed because response options represent a ranking but no clear numerical interpretation, nonparametric 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., percentage of maximum points achieved, mean of scaled items). Then average or sum up the station scores across the OSCE. It is best to calculate station scores only when the station was designed to assess a defined skills-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 across multiple skill domains.
2. Domain scores: For each station calculate subscores (e.g., percentage of maximum points achieved, mean of scaled items) for the items representing specific domains or categories of skill/performance (e.g., com-munication skills, counseling). Then average or sum up the subscores across all stations where a particular domain was assessed.

When designing a blueprint (Step 3, above) 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. As a result, their scores are a more reliable indication of their competence—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 2.15 provides definitions of these key psychometric issues, 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 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—and 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 and OSCEs with fewer stations, Cronbach's alphas should probably exceed 0.60 or 0.70. 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 without intervening training or education) and inter- or intra-rater reliability (comparing checklist ratings among different raters or over time within the same rater).

Once the reliability of a checklist has been established, attention should turn to gathering evidence of its validity, that is, the degree to which it measures what it was intended to measure. There is no simple way to establish validity and instead efforts to support the validity of a checklist should be based on how well it performs: Does it 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

Student Number
(Attach label here)

Evaluator's Name: _____

You have 10 minutes to complete this checklist

OSCE CHECKLIST

COMMUNICATION SKILLS	Not Done	Partly Done	Well Done
Opening			
Introduced self	Did not introduce self **0.7%**	Introduced self only by name and student status **62.4%**	Gave name, student status, and purpose of interview **36.9%**
Information Gathering			
Started with **open-ended** questions	Started with closed, yes/no questions **1.4%**	Began with open-ended questions but stopped prematurely **24.1%**	Started with open-ended questions and continued using them as appropriate **74.5%**
Asked you **what you thought** was the matter	Did not specifically ask **67.8%**	Asked but did not give you enough time to share views **12.1%**	Asked so that you fully shared your views **20.1%**
Managed the **narrative flow** of your story.	Not able to elicit your story because questions not organized logically **6.8%**	Elicited main elements of story, but illogical order of questions disrupted flow **31.1%**	Elicited full story by asking questions that facilitated natural flow of story **62.2%**
Elicited your story using **appropriate questions.**	Impeded story by asking leading questions or more than one question at a time **7.5%**	Used some leading questions and/or asked more than one question at a time, but still able to share most of story **50.0%**	Facilitated the telling of your story by asking questions one at a time without leading you in your responses **42.5%**
Clarified information by repeating to make sure he/she understood you on an ongoing basis	Did not clarify (did not repeat info you provided) **20.1%**	Repeated the information but didn't give you chance to indicate whether accurate **24.8%**	Repeated information and directly invited you to indicate whether accurate **55.0%**
Allowed you to talk **without interrupting**	Interrupted you **13.4%**	Did not interrupt you directly but cut your responses short by not giving you enough time **22.1%**	Did not interrupt and allowed time to express thoughts fully **64.4%**
Relationship Development			
Communicated **concern** or intention to help	Did not communicate intention to help/concern via words or actions **5.4%**	Words OR actions conveyed intention to help/concern **28.4%**	Actions AND words conveyed intention to help/concern **66.2%**
Non-verbal behavior enriched communication (e.g., eye contact, posture)	Non-verbal behavior was negative or interfered with communication **9.5%**	Non-verbal behavior demonstrated attentiveness **50.7%**	Non-verbal behavior facilitated effective communication **39.9%**
Acknowledged your emotions appropriately	Did not acknowledge your emotions **9.5%**	Attempted to acknowledge emotions **51.7%**	Responded to your emotions in ways that made you feel better **38.8%**
Was **accepting/non-judgmental**	Expressed judgment **1.4%**	Did not express judgment but did not demonstrate respect either **28.1%**	Demonstrated respect towards you **70.5%**
Used words you understood and/or explained **jargon**	Jargon made it difficult to understand **4.0%**	Used jargon occasionally but did not significantly interfere with understanding **23.0%**	Provided no opportunity for misunderstanding by avoiding or explaining jargon **73.0%**

Fig. 2.8 Describing OSCE data for a cohort of trainees. Shown for each OSCE checklist item, the distribution of ratings for a class of third year medical students ($n = 160$)

Table 2.15 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 vs. a 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

and peer ratings, etc.)? And ultimately, are checklist scores predictive of actual clinical performance and outcomes?

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 cut-offs to divide learners into those that meet those standards and those that do not. For high stakes examinations, many psychometricians and medical education experts recommend absolute or criterion-referenced cut-offs (i.e., scores that reflect the ability to competently perform specific skills and behaviors). Experts review the "test" (OSCE rating form and cases) content and determine a "passing" score. More complicated methods are also available (Boulet et al. 2003 [review]; Kilminster and Roberts 2004; Krumer et al. 2003).

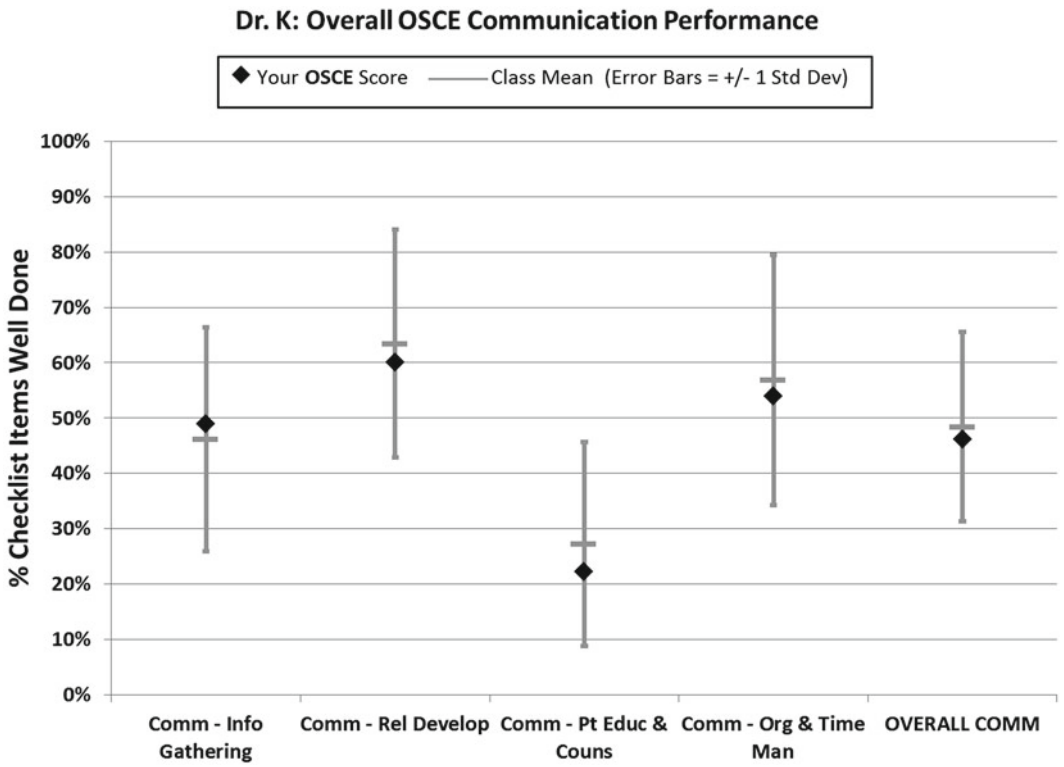

Fig. 2.9 Sample report card illustrating the OSCE performance of an individual learner (following our example General Internal Medicine Residency OSCE). Scores are reported as percentage checklist items "well done" and reflect individual performance across 10 OSCE cases relative to a cohort of 24 OSCE participants (first year residents, in this example)

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 over time (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 (Downing et al. 2006). While exams given on a very large scale can afford—both financially and with respect to numbers of subjects and experts—to go through rigorous standard setting procedures, most smaller-scale projects cannot. Therefore most school or program-based summative OSCEs end up using an approach that combines normative, criterion-based, and practical considerations to setting pass/fail cut-offs.

At NYU we use this combined approach for setting cut-offs to identify students who fail our comprehensive clinical skills exam (CCSE), a summative 8-case 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 as 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 separately, taking what is called a non-compensatory approach (Sadler 2005). We then set a normative passing cut-off at the lowest decile for each competency. Students in this lowest decile across two or more competencies are identified, and then all students who "fail" communication skills alone (because we have found that this is predictive of failure on the USMLE Step II CS exam) are 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 of this book.

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 2.9 provides an example of an OSCE score report. 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 (faculty ratings, exam scores, self-assessments, even clinical and patient data) over time to track and 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, that is, 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.

Best Practices: Managing and Analyzing OSCE Data

- 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.
- Explore the quality of the data in terms of reliability estimates of internal consistency (Cronbach's alpha) before calculating summary scores.
- Calculate OSCE scores based on performance within domains across stations, considering the structure of the data (response options) and how best to derive summaries (percentage well done, average of scaled items, nonparametric methods if necessary).
- Report performance data to learners in ways that are understandable and constructive.
- 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, as a set of cases is created, materials developed, a cadre of SPs recruited, and the team involved gets more experience, organizing OSCEs becomes much easier. By developing a case library such the one exemplified in Fig. 2.10, one can greatly reduce preparations for subsequent OSCEs. It is useful to maintain a library in electronic (backed-up!) and paper format and to make sure that the latest versions of the cases (and training notes) are archived. It is also helpful to maintain a database of SPs and their contact information, and of any evaluative data that may have accumulated for each station. In this way, one can determine whether cases need to be tweaked and whether SPs should be invited back.

Given that licensure exams now include performance-based assessments, and that the ACGME and other accrediting agencies now strongly advocate for the use of OSCEs, it makes sense for organizers to invest energy in institutionalizing OSCEs. Below are some tips for making OSCEs part of the institutional culture.

Case			SP Characteristics		Nature of Encounter			Skills Assessed					
ID	Name	Details	Age	Gender	Acute v Chronic	Presenting Condition	Nature of Visit	Comm	Hx Gather	Phys Exam	Tx Plan & Man	Other	Last Used
1	Work Rounds	Sub-intern and team conduct work rounds	58	F	Acute	Metastatic Breast Ca	Rounds	✓			✓	Pre-cepting	2009
2	Depression	Precept MS IV on ambulatory rotation	64	M	Chronic	Depression	Ongoing	✓	✓		✓	Pre-cepting	2011
3	New Diagnosis	Pt hospitalized while on vacation	55	M	Acute	High BP, Afib	New Visit	✓	✓	✓	✓	System-based practice	2009
54	Diarrhea	Pt calls complaining of diarrhea	35	M	Chronic	Diarrhea	Ongoing	✓	✓		✓	Phone Skills	2010
55	Shoulder Pain	Pt diagnosed with RCC 8 mos ago	62	F	Acute	Metastatic Renal Cell Ca	New Onset	✓	✓	✓	✓		2009
56	Stomach Ache	Pt complaining of stomach pain; worse with GAD	32	F	Chronic	GAD	Ongoing	✓	✓	✓	✓		2011
57	Cholesterol	Pt with hx of HTN and hyperlipemia; noncompliant	45	M	Chronic	HTN, Hyperlipidemia	New Visit	✓	✓		✓		2007
58	Chest Pain	Pt with hx of NIDDM and HTN presents with chest pain	56	F	Acute	NIDDM, HTN	New Visit	✓	✓	✓	✓	Inter-profnl Collab	2010

Fig. 2.10 Snapshot of OSCE cases and characteristics stored in 120-case bank. A database aids in organizing and tracking use of cases and in developing an OSCE blueprint such as that shown in Fig. 2.4

Best Practices: Building Institutional Capacity

- Save all material on an institutional server.
- Create a collaborative interdisciplinary OSCE committee that meets regularly.
- Invite institutional opinion leaders and early adaptors from various departments to observe, help out, stop by.
- Disseminate reports widely.
- Talk about the OSCEs all year round (and with humor!).
- Get the OSCE into your departmental budget.
- Apply for research and program enhancement grants.
- Publish and present experience/findings locally, nationally, and internationally.