Medical education (continued)

Student assessment in the Ohio University College of Osteopathic Medicine CORE system: progress testing and objective structured clinical examinations

RONALD PORTANOVA, PhD; MICHAEL ADELMAN, DO; JOSEPH D. JOLLICK, PhD; STEPHANIE SCHULER, BS; MALCOLM MODRZAKOWSKI, PhD; ELAINE SOPER, PhD; BARBARA ROSS-LEE, DO

It is increasingly apparent that medical education has not kept pace with the evolving healthcare system. In a previous communication, the authors addressed this problem by proposing an “osteopathic prescription for medical education reform.” In keeping with the tenets of that proposal, during the past several years, curriculum reform has been a priority at the Ohio University College of Osteopathic Medicine (OUCOM). Our goal has been to develop a continuum of physician training, that is, a seamless, osteopathic medical curriculum, integrated both vertically (across time) and horizontally (across disciplines), and encompassing the period from matriculation to the completion of a primary care residency. Specific milestones in this endeavor include the following:

1994—Introduction of the patient-centered continuum (PCC), a student-directed, case-based curricular track, founded on continuum education principles, and available to 20% of the undergraduate class;

1995—Establishment of the Centers for Osteopathic Research and Education, or CORE system; this statewide undergraduate and graduate medical education consortium—Osteopathic Postdoctoral Training Institution (OPTI)—provides the administrative and logistical system necessary to deliver a “true” medical education continuum; and

1999—Introduction of the clinical presentation continuum (CPC); this curricular track, inspired by the University of Calgary in Canada, is organized around common patient presentations, largely faculty-directed, incorporates continuum education principles, and is available to 80% of the undergraduate class. (The CPC will be the subject of a forthcoming article.)

One important aspect of medical education reform—some would argue the aspect most important and most difficult to change—is student assessment. In a recent report concerning changes in student assessment implemented by eight schools participating in The Robert Wood Foundation’s “Preparing Physicians for the Future: Program in Medical Education,” Mennin and Kalishman refer to the 30-year history of consistent calls for revision of student evaluation techniques. They conclude that assessment, by its very nature, evokes issues of governance, control, and power and hence is resistant to change. At the same time, they stress the singular importance of assessment in determining the ultimate success or failure of attempts at curriculum reform. A quote by a faculty member at one of the eight participating schools poignantly makes this point: “...[student] assessment does drive the curriculum, and traditional assessment drives the curriculum into tradition. Every place we did not change [assessment], the curriculum is being driven back to whatever it was before....”

At OUCOM, we are inclined to believe that this unidentified faculty member is correct and consequently have included comprehensive, integrated assessment of cognitive and noncognitive performance as explicit components of curricular reform: progress testing and the objective structured clinical examination (OSCE). A status report on our experiences to date, together with a brief review of the pertinent literature, follows.

Progress testing

In medicine and other fields of study, “knowledge” has a central role in determining professional competence. It is therefore imperative that any system of student assessment includes reliable and accurate measures of knowledge acquisition. It is also important to acknowledge the strong influence that the assessment system can potentially have on the learning environment. According to Van Der Vleuten and Verwijnen, an examination system constitutes a set of rules (“behavioral stimuli”) to which students strategically respond to maximize their chance of success. The challenge to meaningful curriculum reform is to break the potential lockstep relationship between testing and learning, that is, to institute a recurring, valid, and meaningful assessment of knowledge mastery and yet avoid having the examination system become the force that drives the students and steers the course of their daily lives, the de facto “coin of the realm.”

In the 1970s, the University of Missouri at Kansas City (UMKC) and

From The Ohio University College of Osteopathic Medicine, and Centers for Osteopathic Research and Education, Athens, Ohio.

Correspondence to Ronald Portanova, PhD, Biomedical Sciences, 320 Grosvenor Hall, Ohio University College of Osteopathic Medicine, Athens, OH 45701.

References

1. Portanova et al. Medical education JAOA • Vol 100 • No 11 • November 2000 • 707
the University of Limburg at Maastricht in the Netherlands\textsuperscript{13} independently devised a method of longitudinal assessment, progress testing, as an approach to this dilemma. More recently, McMaster University in Canada\textsuperscript{14} incorporated progress testing in its undergraduate medical education curriculum. Although the three institutions differ with regard to the specifics of the testing format adopted, all three progress testing systems have the following essential features: (1) the examination is so comprehensive that it is virtually impossible for the student to specifically prepare for the examination, and (2) assessment is based on performance over successive tests and not on the result (passing or failing) of a single examination.

As a working definition, a progress test is a method for assessing the time-dependent acquisition and retention of knowledge, relative to curriculum-wide goals and end objectives, that is, a method of assessing the accumulated knowledge of trainees, relative to the expected (required) knowledge domain of a prototypical “finished product,” that is, the “ideal” graduate of the training program.

**Progress testing in the OUCOM-CORE system: continuum progress indicator**

In November 1999, after several years of deliberation, the OUCOM-CORE system inaugurated a pilot progress testing program—the continuum progress indicator (CPI). In making this decision, we were inspired by the pioneering work at UMKC, the University of Limburg, and McMaster University (cited previously) and by a workshop on progress testing conducted at the Generalists in Medical Education Meeting in October 1994.\textsuperscript{15} The format and administrative framework of the examination that we adopted is closely modeled on that used by UMKC.

**Characteristics of the continuum progress indicator**—The following are characteristics of the CPI:

- The CPI assesses knowledge, not skills. The knowledge tested may be factual, procedural, and/or germane to attitudes. At present, neither the CPI nor other progress tests place an emphasis on assessing reasoning ability (pure, applied) and clinical problem-solving skills; however, we are moving in this direction by modifying the test bank to include more questions of this type.
- The examination is a tool for gauging progress (advancement) toward completion of the training program (the completion of primary care residency training) and is not intended as a means of certifying educational achievement relative to steps along the way or of the finished product (as do end-of-course, “final” examinations and “board” examinations, respectively).
- The knowledge domain assessed is dictated by curricular objectives and content over the entire 7-year training program; that is, the accumulated knowledge of trainees is assessed relative to the knowledge domain of an ideal primary care resident on the last day of the training program.
- Each sitting of the 6-hour examination consists of 330 multiple-choice (five choices) questions, distributed by discipline, approximately as follows:
  - 30% basic medical science, including anatomy (gross, micro), biochemistry, immunology, microbiology, pharmacology, physiology, and psychosocial.
  - 30% medicine, including infectious, cardiovascular, metabolic, musculoskeletal, gastrointestinal, genitourinary and respiratory disease; dermatology; endocrinology; neurology; hematology; oncology; and toxicology.
  - 10% general surgery, distributed by organ-system (for example, gastrointestinal) and including emergency medicine and trauma.
  - 10% pediatrics, distributed as for internal medicine and including growth, development, and child behavior.
  - 10% obstetrics, including pre- and postnatal care, labor, and delivery; and gynecology, including infectious diseases, neoplasms, fertility, and contraception.
  - 10% osteopathic principles and practices, distributed according to the five aforementioned disciplines.
- Students at all levels of the training program are tested repeatedly as they advance through the curriculum. The CPI was initiated in November 1999 and administered for the second time in April 2000. The examinees consisted of students enrolled in OUCOM’s PCC, a problem-based curricular track, and a selected group of interns and residents enrolled in CORE training programs. The results from the first two sittings of the CPI are described below. The CPI will continue to be given twice annually (fall and spring); it will be extended to all undergraduate medical students (both of our curricular tracks) and made available to all postgraduate trainees in the CORE system.
- A different set of test items (questions) is used for each sitting of the examination; however, it must be stressed that for any given sitting, trainees at all levels of the 7-year program take an identical examination.
- As CPI data are accumulated, they will be evaluated for potential correlates with admissions (student selection) criteria, validity (including predictive), and reliability.

**Rationale for introducing the CPI and expected benefits**—The CPI was introduced with the understanding that, pending validation as described above, it would be used only as a means of formative self-assessment by the student and as an adjunct to program evaluation. Individual examination results are returned to examinees, who are also provided with group scores and a comparison of how they performed in relation to peers and trainees at other levels of the curriculum; faculty and program administrators receive group, but not individual, scores. Even in this narrow context, it is expected that progress assessment offers a number of benefits to the primary stakeholders. In the case of students, the examination provides the following:
- An opportunity for identifying strengths and weaknesses, overall and in various disciplines, in a nonthreatening environment; potential problem areas can be identified and addressed prospectively.
- Reduction in anxiety concerning whether the student is making adequate
progress in the breadth and depth of coverage in their studies; the latter is especially significant in a curriculum like OU COM’s PCC, where students are empowered to select their own learning issues.

- Direction and assurance concerning programmatic end objectives (undergraduate and graduate), graduation criteria, and certifying examinations.
- An exercise in adult learning and continuing education, including the development of metacognitive skills.

For faculty, program directors, and administrators, the examination provides these features:
- A means of evaluating students and programs that is consistent with continuum education principles; narrow course-specific objectives are deemphasized in favor of integrated medical-scientific end goals and objectives.
- Pending validation, a timely, reliable basis for addressing deficiencies (in individuals) and for comparing the strengths and weaknesses of different programs (for example, postgraduate training programs at different sites).
- A means to assess the competence level of students transferring into or reentering the educational program after an absence.
- Rich source of data for curricular research and program development and evaluation.

**Results and conclusions—** Table 1 summarizes the results obtained on the first two administrations of the CPI: November 1999 and April 2000. The data presented are mean percent-correct scores plus or minus one standard deviation for year-1 to year-4 undergraduate students (Y1 to Y4) and for a mixed group (interns and residents) of postgraduate trainees; N = number of individuals in each group. Scores are reported for each of the disciplines tested (BS, biomedical science; MED, internal medicine; OBG, obstetrics and gynecology; SUR, surgery; PED, pediatrics; and OPP, osteopathic principles and practice); overall performance on the examination is also indicated. As expected, the data indicate that overall performance, and performance in most of the subdisciplines, varies directly, but tends to level off with the level of training (compare Y1 to PG for each of the examinations). For a given group (class), performance tends to increase as

### Table 1

*Results of First Two Continuum Progress Indicator Examinations by Year and Discipline*

<table>
<thead>
<tr>
<th>Discipline</th>
<th>November 1999</th>
<th>April 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Post-graduate</td>
<td>Year 4</td>
</tr>
<tr>
<td></td>
<td>(N=17)</td>
<td>(N=12)</td>
</tr>
<tr>
<td>Basic science</td>
<td>62.3±8.8</td>
<td>65.6±10.2</td>
</tr>
<tr>
<td>Medicine</td>
<td>77.4±6.0</td>
<td>75.0±7.6</td>
</tr>
<tr>
<td>Obstetrics and gynecology</td>
<td>83.8±5.4</td>
<td>79.2±6.6</td>
</tr>
<tr>
<td>Surgery</td>
<td>80.6±5.9</td>
<td>79.6±7.9</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>83.8±5.8</td>
<td>77.6±8.2</td>
</tr>
<tr>
<td>Osteopathic principles and practice</td>
<td>64.1±14.6</td>
<td>66.7±6.5</td>
</tr>
<tr>
<td>Overall</td>
<td>74.8±3.0</td>
<td>73.6±6.5</td>
</tr>
<tr>
<td></td>
<td>Post-graduate</td>
<td>Year 4</td>
</tr>
<tr>
<td></td>
<td>(N=127)</td>
<td>(N=14)</td>
</tr>
<tr>
<td>Basic science</td>
<td>65.2±6.4</td>
<td>74.3±6.9</td>
</tr>
<tr>
<td>Medicine</td>
<td>75.3±4.9</td>
<td>77.2±8.3</td>
</tr>
<tr>
<td>Obstetrics and gynecology</td>
<td>69.5±6.6</td>
<td>68.8±3.6</td>
</tr>
<tr>
<td>Surgery</td>
<td>71.7±7.3</td>
<td>79.9±4.6</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>72.7±7.0</td>
<td>78.3±4.3</td>
</tr>
<tr>
<td>Osteopathic principles and practice</td>
<td>51.3±1.8</td>
<td>48.3±2.0</td>
</tr>
<tr>
<td>Overall</td>
<td>71.4±6.6</td>
<td>73.1±9.1</td>
</tr>
</tbody>
</table>

Portanova et al • Medical education  
JAOA • Vol 100 • No 11 • November 2000 • 709
students advance through the curriculum. For example, first-year (Y1) students scored 47.2% on basic sciences (BS), 40.3% on surgery (SUR), and 43.6% overall in November 1999, and 55.8%, 52.8%, and 49.4%, respectively, in April 2000.

These findings are consistent with those of other institutions employing progress testing, taking into account that previously reported data are largely restricted to undergraduate medical education, that is, does not span the period from year 1 of medical school through postgraduate training, as does the CPI. At present, we are encouraged to continue to deliver the CPI twice annually (fall and spring), to extend it to all of our undergraduate students (both of our curricular tracks), and to make it available to all postgraduate trainees in the CORE system.

Objective structured clinical examinations
The OSCE grew out of an increasing dissatisfaction with the limited capabilities of traditional oral and written examinations to test for clinical skills. In contrast to knowledge-based assessment, which allows for a broad sampling of student’s cognitive database (that is, progress examinations), performance-based assessment should test what the learner “does” rather than just what he or she “knows.”

Introduced more than 25 years ago by Harden and colleagues, an OSCE involves the rotation of examinees around a circuit of stations at which they are required to perform a variety of clinical tasks. These tasks may include history taking; performing portions of a physical examination; undertaking technical procedures such as suturing; ordering or interpreting diagnostic studies; counseling patients; or handling an emergency trauma victim. Use of clinically relevant tasks tailored to the skills to be assessed; controlled and standardized testing situations; and predefined grading criteria provide a significant advance over traditional, unstructured, preceptor-rated performance evaluations in clinical training.

While individual OSCE stations may use any type of assessment method, ranging from observation of procedural skills to answering postencounter questions, Van Der Vleuten noted that, most frequently, standardized patients are used to test hands-on clinical skills. "Standardized patients" may be nonphysicians who are taught to portray patients in a standardized and consistent fashion. These “patients” might simulate multiple physical findings such as bruises, abnormal reflexes, diminished breath sounds, depression, or signs of substance abuse. Examinees interact with the standardized patients as though they were examining and counseling real patients.

All examinees see the same clinical problems portrayed by the standardized patients. The usual OSCE station encounter lasts 5 minutes; however, OSCEs may employ 5- to 20-minute stations depending on the particular clinical skills being evaluated. Examinees are rated on detailed checklists by the standardized patient, by an evaluator in the examination room, or by both.

As with the progress examination, it should be virtually impossible for students to specifically prepare or “cram” for the OSCE. Rather, the OSCE should be an assessment of skills learned over a specified time period, such as the completion of specific clinical rotations. In addition, the OSCE should also be used to assess the skills learned throughout the continuum of the student’s medical education, including residency training.

Objective structured clinical examinations in the OUCOM-CORE system
As noted above, in 1995, OUCOM and thirteen of its training hospitals throughout Ohio developed an educational consortium—CORE—with the primary goal of strengthening both the predoctoral and the postdoctoral training delivered at the hospital sites.

OUCOM students complete the last 21 months of their medical education at one of the 13 hospital sites within the CORE system. The CORE was structured to provide the same high level of training to all students, irrespective of the hospital site where they were based. However, as do most medical schools, OUCOM relies heavily on preceptor evaluation and assessment of students to determine student progress. Under this method of assessment, standardization of preceptor evaluation of students throughout the state was difficult to achieve. The same student taking the same rotation at two different CORE hospitals might get different evaluations because of preceptor variability and evaluation techniques.

Many schools add pre- and postrotation examinations to better quantify a student’s knowledge base. OUCOM also uses this evaluation tool on some rotations. Additionally, the implementation of a progress examination, the CPI, also helps to quantify the student’s level of knowledge and understanding. However, while these tools examine a student’s factual knowledge, they fail to adequately evaluate clinical skills, interpersonal skills, and student professionalism. The OSCE gives us the opportunity to assess these skills in a CORE-wide, reliable manner.

Three years ago, a pilot OSCE was initiated for the fourth-year medical students located at one of the five regions of the OUCOM-CORE system. A needs assessment was performed to determine the skills that we wanted the OSCE to evaluate. Using these data, twelve stations were developed. These included one station for each of the required student rotations—surgery, orthopedics, psychiatry, family medicine, internal medicine, gynecology, emergency medicine, and pediatrics, as well as stations on medical ethics, dermatology, radiology, and osteopathic principles and practices. Case development was led by our OSCE team with invaluable assistance from specialists in each area and focused on history and physical examination skills, procedures, physician-patient relationships, and ethical versus legal issues.

Over the next 2 years, the cases and the evaluation tool were refined, and in November 1999, our first statewide OSCE was administered to all of the OUCOM fourth-year students in all five CORE regions. (A subsequent report will...
detail the process and budgetary issues that we encountered in the development of a statewide program.)

Characteristics of the OSCE—The OSCE assesses clinical skills, knowledge, and professionalism. As with the CPI, the OSCE is a tool for gauging progress toward completion of the training program. In this case, it focuses on the everyday skills a physician uses in practice. The OSCE is not used for formal grading purposes or to determine advancement. It serves as a window of opportunity to observe the strengths and weaknesses of the students as well as of the training programs themselves. The data collected will assist in demonstrating the consistency or lack thereof in medical education among the various teaching sites within the consortium.

Rationale for introducing the OSCE and expected benefits—The OSCE was initially introduced into the OUCOM-CORE system as a tool to evaluate student skills that could not be adequately evaluated by written or oral examinations. However, as we soon came to understand, the OSCE also opened up many other evaluative opportunities. First, it gave students an opportunity for self-evaluation. Students constantly question their clinical skills; the OSCE gives them the ability to assess their strengths and their standing, clinically, as compared to their peers. Second, preceptors acted as evaluators for the OSCE and thus had an opportunity to assess what students actually learned during their rotations. This led to many of the preceptors modifying their teaching techniques and learning activities, thus improving the overall rotations. Finally,
it gave us an opportunity to compare clinical rotations at all of our regional sites. Through the data collected, we identified rotations needing more attention and those that functioned well. By focusing on the more problematic rotations, we will be able to improve student learning. This will further lead us to our ultimate goal of making sure that no matter where our students are trained within the consortium, they will all receive the same high-quality education.

**Results**—Table 2 summarizes the results obtained in our first statewide OSCE. For each of the individual hospitals (Hospital 1 to Hospital 14), the data presented are mean correct responses at each station (ethics, acute abdomen, etc), maximum points available at each station, and the CORE-wide average and descriptive statistics for the station. The data are also grouped by regions of the state (COREs) as follows: NW, northwest; NE, northeast; C, central; SW, southwest; and SE, southeast. While the data indicate rather large differences between group means among the COREs and hospitals, based on an analysis of variance, the differences were not statistically significant. This may be due to the fact that the OSCEs were given to existing groups of students rather than being based on a random selection of equal numbers of students at each training site. Group size varied from 4 students at Hospital 2 to 12 students at Hospital 1 and from 10 students in CORE-C to 32 students in CORE-NE.

Despite the lack of statistical significance, we identified areas of improvement for student performance as well as for individual programs throughout the consortium. A second statewide fourth-year OSCE is planned for November 2000. In addition, a pilot for a family medicine residency OSCE was completed in May 2000, and a full statewide family medicine OSCE was completed in August 2000. Plans are also under way for consortium-wide OSCEs for interns, emergency medicine residents, pediatrics, and surgery.

**Comment**
Individually, the progress examination and the OSCE are excellent evaluation tools. Further, the combination of the progress examination, which assesses the accumulated knowledge of trainees, and the OSCE, which assesses the accumulated clinical and interpersonal skills of trainees, provides an excellent method of comprehensively evaluating medical students, interns, and residents relative to the expected (required) skills and knowledge domain of a prototypical, “ideal” graduate of the training program.

**Acknowledgment**
The authors thank T. Lee Willoughby, MS, and Louise Arnold, PhD, of the University of Missouri at Kansas City for their assistance in preparing the continuum progress indicator used at the Ohio University College of Osteopathic Medicine.

**References**