

# TOPICS IN TRAINING

## Learning Assessment Toolkit

By Peter G. de Boer, MA, FRCS, Rick Buckley, MD, FRCS(C), Pascal Schmidt, Robert Fox, EdD, and Jesse Jupiter, MD

Modern surgical education is evolving with changes in the academic environment. Surgeons, surgical educators, and administrative associations need to know what to teach, how to do it, and whether it has succeeded. Whereas surgeons have many tools to help them to see what is needed for better patient care and to measure patient progress, surgical education is just beginning to develop techniques of interpreting the needs of students and to assess outcomes of learning in their practice<sup>1,2</sup>. The absence of surgical education assessment techniques has made it necessary to guess what is needed to improve education. We have estimated the success or failure of teaching by attempting to score faculty performance rather than on obtaining more valid data. This has led surgical educators to use mistaken assumptions based on their personal experience or reasoning without evidence<sup>3</sup>.

Surgical education should be based on needs assessment, efficient program planning, and a strong curriculum. The motivation of learners should be high, and the learners' needs should be met. This is dependent on measurable outcomes with objective data demonstrating knowledge acquisition that meets the needs of surgical learners<sup>4</sup>.

In the last ten years, the providers of educational resources—governments, training boards, charitable foundations, and commercial companies—have been increasingly interested in whether sur-

gical education, as it is delivered, has had a measurable effect<sup>5-7</sup>.

Through a series of pilot-tested steps, we developed a set of instruments, providing insights into the effectiveness of surgical education in the field of orthopaedic trauma. We termed the instrument the Learning Assessment Toolkit. It was developed to supplement the judgment of surgical educators before and after a teaching event with real evidence of need, motivation, and outcomes of educational programs in orthopaedic trauma surgery.

The primary goal of this report is to outline the elements of the Learning Assessment Toolkit and its developmental steps. A secondary goal is to show how its use can change the nature and content of surgical educational events to improve learning outcomes. This report also suggests what further instruments are necessary to achieve the desired end result: education that substantially improves patient care.

### Materials and Methods

The educational event used to design and test the assessment toolkit is a fracture course based on the AO principles of operative fracture management. The course is aimed at residents in their first few years of training. The participant groups are homogeneous in North America and Western Europe but heterogeneous in the developing world in terms of their needs and experience. The

course has evolved over the past forty-nine years, but it has been taught in a standardized form for the past decade.

The “key competencies” guide the development of the Learning Assessment Toolkit, and the teaching and learning in the course (Table I). A *key competency* was defined as a piece of knowledge and/or skill that educators expected the course participants must know or be able to do after the course<sup>8,9</sup>. Key competencies are statements describing what behaviors are necessary to address problems related to successfully providing the gold standard for patient care. These responses were collected from three experienced course chairmen, all with greater than ten years of teaching experience acting as a panel of experts.

### How the Learning Assessment Toolkit Works (Table II)

#### Precourse Assessment

Two weeks before the educational event, course participants were contacted online. They were presented with the fourteen key competencies and were asked two questions about each competency. The first question was: “How important is this competency to you in your daily practice?” Participants were asked to score from 1 to 5 (with 5 indicating the highest importance) on a Likert scale. They were then asked to evaluate their own ability relating to the individual competencies using the same scoring system.

**Disclosure:** The authors did not receive any outside funding or grants in support of their research for or preparation of this work. Neither they nor a member of their immediate families received payments or other benefits or a commitment or agreement to provide such benefits from a commercial entity.

**TABLE I Competencies Defined for Principles of Fracture Management Course**

- 1 Assess fracture-related soft-tissue injuries and apply strategies to treat these injuries
- 2 Safely treat open fractures
- 3 Demonstrate techniques to achieve relative stability based on an accurate description of indications
- 4 Diagnose and treat compartment syndrome
- 5 Explain how design features and functions of various implants relate to their application to fracture fixation
- 6 Accurately explain the influence of reduction and fixation techniques on bone healing
- 7 Apply reduction techniques in fracture management
- 8 Apply techniques to stabilize the pelvis in the emergency management of a hemodynamically unstable patient with a pelvic fracture
- 9 Explain the factors that lead to malunion, nonunion, and infection following fracture treatment
- 10 Explain why the components of preoperative planning are essential for successful fracture treatment
- 11 Treat diaphyseal fractures according to the principles of restoring length, axis, and rotation
- 12 Treat articular fractures to achieve early active movement according to the principles of anatomical reduction and absolute stability
- 13 Demonstrate techniques to achieve absolute stability based on an accurate description of indications
- 14 Explain strategies for treating fractures in osteoporotic bone

Three pieces of information were obtained from this survey, passed on to the faculty of the course, and fed back to the individual course participants.

1. Which competencies were rated as being the more important from the participants' point of view—an indication of where they thought they ought to be in their practice?

2. How capable did the participants believe they actually were in each competency?

3. What was the difference between these two measures for each key competency, i.e., the gap score? This is an indication of the difference between where the participants thought they were and where they ought to be in their practice.

The gap score is a reasonable measure of the motivation of the course participants to learn at the course. Discomfort over the perception of where adult learners believe they are and where they believe they ought to be, as indicated in this case by the measure of the importance of a given competency, creates discomfort in the learner and spurs the drive to learn and change<sup>10</sup>.

At the same time that the course participants were contacted online with regard to their needs assessment, they were also given two multiple-choice questions relating to each key competency. The multiple-choice test questions were developed by the surgical educators through an expert panel<sup>11</sup> consisting of existing experienced faculty members and a small group of senior North American orthopaedic surgical residents following best practice guidelines of multiple-choice question-writing (Fig. 1).

The test questions were placed into assessment software (Questionmark Perception; Questionmark, Norwalk, Connecticut) for online pilot testing to collect statistical data as to how the learners were answering the individual questions. After pilot testing these questions in three courses, the expert panel was reassembled to study the data obtained and to eliminate or refine questions that were too easy, too hard, or too confusing. Following further pilot testing of the adapted questions, they were reviewed again and became part of a library of test questions.

The response patterns for each pilot have been continually assessed to build up evidence as to the validity and discriminatory capacity of each individual test item.

The objective assessment enables the faculty to have insight into whether the course participants' assessment of their current performance was or was not accurate and enables learners to have an understanding as to their true level of knowledge. Previous studies have shown that a doctor's perception of his or her own level of knowledge or skill is not accurate, with a tendency for doctors to overestimate their own abilities<sup>12-14</sup>.

#### Evaluation During the Course

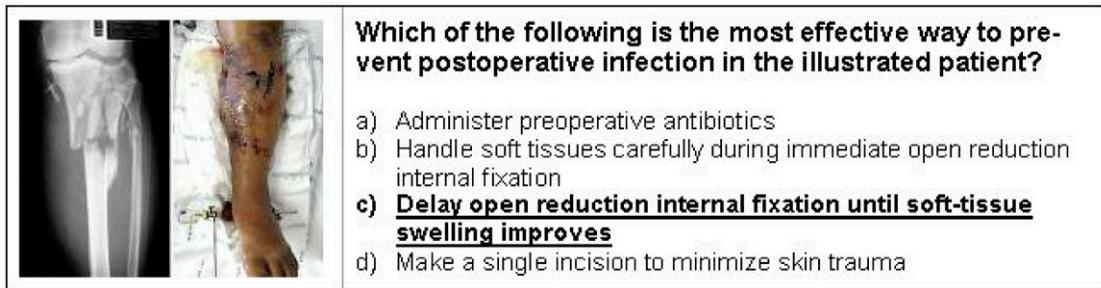
The course participants were asked to evaluate each presentation—lecture, discussion group, or practical skill session. The evaluation system for each presentation was on a 5-point Likert scale with use of an audience response system. Two questions were asked:

1. How relevant is the presentation to your daily practice?

2. How effectively was the presentation given?

**TABLE II Components of Learning Assessment Toolkit**

Precourse
Subjective needs assessment (self-assessment of key competencies)
Objective needs assessment (test of knowledge with multiple-choice questions)
Course
Participant evaluation of lecture, practical, or discussion
Relevance
Faculty performance
Faculty evaluation of presentation (evaluation based on objective criteria)
Postcourse
Subjective needs assessment (self-assessment: how much did the participants think they had learned?)
Objective needs assessment (test with multiple-choice questions to show how much had been learned)



**Which of the following is the most effective way to prevent postoperative infection in the illustrated patient?**

- Administer preoperative antibiotics
- Handle soft tissues carefully during immediate open reduction internal fixation
- Delay open reduction internal fixation until soft-tissue swelling improves
- Make a single incision to minimize skin trauma

Fig. 1

A sample question from the precourse and postcourse objective assessment used to assess Competency 1: Assess fracture-related soft-tissue injuries and apply strategies to treat these injuries.

The course participant rating was collected electronically after each session, with participation rates in excess of 80%. To ensure validity as to whether the course participants' assessment of faculty performance was or was not accurate, two faculty members were also assigned to assess each presentation on the basis of a group of ten criteria that had been agreed on by the faculty before the educational event and that were supported by the available literature<sup>15-17</sup>.

#### Postcourse Assessment

Two weeks after the course, the participants were contacted online. They were asked to repeat the online questionnaire. They were also given two multiple-choice questions for each competency. The set of questions was new to each individual participant but had been asked before the course to the other half of the course participants. Four questions (A, B, C, and D) were allocated to each course. Half of the course participants answered questions A and B before the course and answered C and D afterward. Half of the participants answered questions C and D before the course and answered A and B afterward. Participants were not asked the same questions before and after the course to avoid test-retest bias of their answers, which would have tended to improve their scores after the course, giving a false impression about knowledge transfer occurring as a result of the course.

#### Source of Funding

This project was funded entirely by the AO Foundation, Davos, Switzerland.

#### Results

The Learning Assessment Toolkit has been used in twenty courses in eight countries involving 1812 participants originating from forty-seven countries. However, only data collected from courses that took place after finalization of the assessment questions are presented—eleven courses in six countries involving 912 participants from forty-six countries. Response rates ranged from 41% to 98%, with an average of 62% for the assessments before the course and 51% for the assessments after the course.

#### Precourse Subjective Needs Assessment

Overall gap scores were large for an educational event<sup>10</sup>. The average gap score was 2.25, with a fairly narrow range between 1.9 and 2.4. Certain competencies were consistently ranked as being more important than others, and this pattern was independent of the geographical location of the course. The "emergency management of a hemodynamically unstable patient with a pelvic fracture" was consistently identified as the highest need by the course participants.

Large gap scores can occur in one of two ways—either the participant ranks the competency to be very important or the participant believes that his or her ability is poor for the competency tested. Similarly, a small gap score can be explained in two ways—either the participant thinks that the competency is not important for his or her practice or that his or her ability is reasonable for the competency tested. Those competencies, which consistently showed large and small gap scores, are listed in Table III.

#### Precourse Knowledge Assessment

The level of knowledge of the course participants varied from course to course. On the average, the questions were answered correctly by 59% (range, 51% to 67%) of the course participants. The courses held in the developing world were attended by surgeons with greater experience, and this was reflected in their higher knowledge scores.

The correlation between knowledge, as measured by objective testing and by self-assessment, was variable. In three courses, participants who rated their need to learn about compartment syndrome as low because they believed they had good existing abilities with regard to that subject were incorrect in their self-assessment as they had low scores on their objective assessment. In two courses, participants rated their need to learn about compartment syndrome as low because of good existing abilities, and objective testing showed them to be correct.

#### Course Evaluation

Electronic evaluation of faculty performance by the course participants was carried out in four courses. The other courses were evaluated with use of a paper-based system. A total of 45,600 responses were analyzed. There was a very strong correlation between the participants' perception of the relevance of the presentation to their practice (average score, 4.04; range, 3.88 to 4.21) and to their perception of faculty performance (average score, 3.99; range, 3.77 to 4.17). When presentations were given about the same subject to different audiences by different faculty, there was a wide variation in the participants' assessment of

TABLE III High and Low Gap Scores\*

Large Gap		Small Gap	
Large Gap Score: High Level of Importance	Large Gap Score: Low Level of Knowledge	Small Gap Score: Low Level of Importance	Small Gap Score: High Level of Knowledge
Emergency management of a hemodynamically unstable patient with a pelvic fracture	Treatment of articular fractures	Principles of preoperative planning	Diagnose and treat a compartment syndrome
How design and function of implants relate to their application in fracture fixation	Techniques of relative stability		
Techniques of absolute stability			

\*Gap scores are a measure of the difference between where a learner wants to be and where the learner thinks he or she is. Gap scores are derived from self-assessment and are purely subjective.

performance. The participants' perception of relevance also changed following the change in performance perception. Therefore, basing changes in curriculum on an analysis of perception of relevance in isolation from faculty performance may be invalid.

The evaluation of faculty performance by a trained faculty assessor on the basis of agreed-on criteria correlated very weakly with the participants' evaluations of either relevance or performance (Pearson correlation coefficient,  $r = 0.54$ ;  $p < 0.0001$ ). On those occasions when the performance and relevance scores were not closely related, the faculty assessment closely followed the gap between the two different evaluation criteria. For example, a presentation ranked very effective by the faculty evaluator was very likely to have a performance score considerably higher than the relevance score.

#### Postcourse Subjective Evaluation

All courses evaluated showed marked decreases in the gap scores measured two weeks after the course (average, 1.17; range, 0.55 to 1.645). The gap scores of all competencies declined, with the biggest decreases occurring in the competencies that had had the highest needs before the course. These

figures reflect the belief that the course participants believed that they had learned as a result of the course. The learners' highest residual needs varied from course to course, but the "management of open fractures" and "the emergency management of a hemodynamically unstable patient with a pelvic fracture" were the most common areas of residual perceived need.

#### Postcourse Objective Assessment

Most competencies showed an improvement in their objective assessment scores (average, 73%; range, 69% to 77%). There was considerable variation in learning outcomes for each of the competencies. Problem areas were in the teaching of reduction techniques, preoperative planning, and compartment syndrome, where improvements in learning outcomes were modest.

#### Discussion

With accreditation changes, continuing education for surgeons must meet new requirements<sup>18</sup>. In addition to having well-prepared faculty who present information in a thoughtful and organized manner, new standards for the accreditation of continuing medical education specify that programs must be based on learner gaps in knowledge, performance,

or patient health status. Course administrators must document the assessment of these outcomes<sup>18</sup>. This presents a formidable challenge because tools for assessing needs, motivation, and outcomes in terms of gaps in knowledge and performance have not been available. The Learning Assessment Toolkit provides a short practical system for discovering objective and self-assessed gaps in performance of key competencies before and after educational programs.

The toolkit data are designed to provide accurate information related to level of competency for surgeon performance by using case-based multiple-choice questions to test clinical judgment and decision making. Objective evidence and perception together provide feedback to the learners and teachers with regard to the learners' level of motivation and their gaps in knowledge and skill before and after a learning experience. With these kinds of data, educators can understand the level of motivation before and after instruction and also assess gaps in knowledge and skill related to solving clinical problems before and after instruction. The educator learns how learners perceive themselves, how accurate these perceptions are, and to what extent an educational activity has changed perceptions and actual knowledge.

This information can help learners to correct their self-assessed weaknesses and guide them in self-directed learning activities. After the educational experience, learners are given personal data with regard to their perceptions and their individual scores on objective questions related to clinical cases. This improves the accuracy of their self-assessment and can help them to plan for future participation in continuing medical education events.

The education of doctors has one major purpose: to produce changes in knowledge that result in improved patient care. The Learning Assessment Toolkit provides objective evidence as to the success or failure of an educational event in producing improved levels of knowledge<sup>19</sup>. It provides information to educators as to the strengths and weaknesses of their program and

provides evidence on which effective future changes can be made.

This evaluation system provides a useful guide to enable educators to design appropriate educational offerings to meet the needs of surgeons. However, if individuals learn knowledge and skills in an education event but cannot put their skills into practice after the event, then the event clearly has not been successful. Assessment of the barriers to knowledge implementation after a course is therefore critical. We are presently conducting a study to identify barriers that are encountered by doctors in implementing what they have learned at courses and how educators can help them to overcome these barriers<sup>20</sup>.

A Learning Assessment Toolkit can be used by any group of educators.

All that is required is the creation of a list of competencies for the educational event and the test questions designed to test knowledge of these competencies. Many software packages exist to allow online testing. The assessment toolkit described in this paper is not subject to copyright, and the authors would be pleased to assist any groups interested in setting up their own assessment program.

NOTE: The authors thank Dr. Laurent Audigé, DVM, PhD, Manager, Methodology, AO Clinical Investigation and Documentation, for his invaluable help in the statistical analysis within this paper.

Peter G. de Boer, MA, FRCS  
Pascal Schmidt  
AO Education, Stettbachstrasse 6,  
CH-8600 Dübendorf, Switzerland.

E-mail address for P.G. de Boer:  
Piet.deboer@aofoundation.org.  
E-mail address for P. Schmidt:  
Pascal.schmidt@aofoundation.org

Rick Buckley, MD, FRCS(C)  
Foothills Medical Centre, AC 144A,  
1403-29th Street NW, Calgary,  
Alberta T2N 2T9, Canada.  
E-mail address: buckclin@ucalgary.ca

Robert Fox, EdD  
1001 Olde Oak Court,  
Norman, OK 73026.  
E-mail address: drrdfox@me.com

Jesse Jupiter, MD  
Department of Orthopaedic Surgery,  
Massachusetts General Hospital,  
Yawkey Building 2100,  
55 Parkman Street,  
Boston, MA 02114.  
E-mail address: JJupiter1@partners.org

## References

- Spencer JA, Jordan RK. Learner centered approaches in medical education. *BMJ*. 1999;318:1280-3.
- Davis DA, Barnes BE, Fox RD. The continuing professional development of physicians: from research to practice. Chicago: AMA Press; 2003.
- Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. *JAMA*. 1995;274:700-5.
- Wilkes M, Bligh J. Evaluating educational interventions. *BMJ*. 1999;318:1269-72.
- Casebeer L, Raichle L, Kristofco R, Carillo A. Cost-benefit analysis: review of an evaluation methodology for measuring return on investment in continuing education. *J Contin Educ Health Prof*. 1997;17:224-7.
- Regnier K, Kopelow M, Lane D, Alden E. Accreditation for learning and change: quality and improvement as the outcome. *J Contin Educ Health Prof*. 2005;25:174-82.
- Bennett NL, Davis DA, Easterling WE Jr, Friedmann P, Green JS, Koeppen BM, Mazmanian PE, Waxman HS. Continuing medical education: a new vision of the professional development of physicians. *Acad Med*. 2000;75:1167-72.
- Norman GR, Davis DA, Lamb S, Hanna E, Caulford P, Kaigas T. Competency assessment of primary care physicians as part of a peer review program. *JAMA*. 1993;270:1046-51.
- Ward J, Macfarlane S. Needs assessment in continuing medical education. Its feasibility and value in a seminar about skin cancer for general practitioners. *Med J Aust*. 1993;159:20-3.
- Fox RD, Miner C. Motivation and the facilitation of change, learning, and participation in educational programs for health professionals. *J Contin Educ Prof*. 1999;19:132-41.
- Ebel RL. Essentials of educational measurement. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall; 1979.
- Ward M, MacRae H, Schlachta C, Mamazza J, Poulin E, Reznick R, Regehr G. Resident self-assessment of operative performance. *Am J Surg*. 2003;185:521-4.
- Dunning D, Heath C, Suls JM. Flawed self-assessment. Implications for health, education and the workplace. *Psychological Science in the Public Interest*. 2004;5:69-106.
- Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296:1094-102.
- Green JS, de Boer PG. AO principles of teaching and learning. Stuttgart: Georg Thieme Verlag; 2005.
- Coats W, Smidchens U. Audience recall as a function of speaker dynamism. *J Educ Psychology*. 1966;57:189-91.
- Russell IJ, Hendricson WD, Herbert RJ. Effects of lecture information density on medical student achievement. *J Med Educ*. 1984;59:881-9.
- ACCME. <http://www.accme.org>. Accessed 2010 Jan.
- Sachdeva AK. The new paradigm of continuing education in surgery. *Arch Surg*. 2005;140:264-9.
- Mazmanian PE, Daffron SR, Johnson RE, Davis DA, Kantrowitz MP. Information about barriers to planned change: a randomized controlled trial involving continuing medical education lectures and commitment to change. *Acad Med*. 1998;73:882-6.