

Designing and Implementing a Harmonized Evaluation and Assessment System for Educational Events Worldwide

Monica Ghidinelli, PhD, Michael Cunningham, PhD, Miriam Uhlmann, PhD, and Alain Rickli, MA, AO Education Platform, and Urs Rüetschi

Summary: To determine the effectiveness of educational events and improve the quality of continuing medical education (CME), course providers and medical faculty instructors must have access to structured and consistent collection and reporting of evaluation and assessment data. In 2012, the AO Foundation (Switzerland) used a wide range of evaluation questions and processes that were inconsistent across various clinical areas. With AO's 700 educational events delivered annually in multiple geographical regions, it was therefore challenging to determine overall education effectiveness and to identify and compare trends and topics based on individual course data. This led to a decision by AO to update, align, and harmonize the various questions and processes to create a new streamlined and consistent evaluation and assessment system. A series of expert advisory group sessions and consensus meetings were convened over a 3-year period, and feedback from 8 stakeholder groups was incorporated. AO developed processes and online tools that were piloted in several educational events and then implemented worldwide. Faculty and course organizers were trained to gather and apply the information. In 2019, this new course evaluation and assessment system was applied to more than 70% of AO's yearly educational events. The generated reports have helped faculty to adjust educational events to meet the needs of participants. The new system has also helped committees and regions to plan future educational events and to improve the quality of CME on an ongoing basis.

Key Words: evaluation, assessment, CPD, CME, evaluation design
(*J Orthop Trauma* 2021;S5–S10)

INTRODUCTION

Continuing medical education (CME) and continuing professional development (CPD) are designed to help participants to close knowledge gaps or improve clinical

performance for better patient outcomes.¹ To determine if an educational activity achieves these goals, information from the collection of evaluation and assessment data is essential. This information also guides course providers when making decisions and taking actions regarding future educational activities.²

Evaluation is a process of observing and measuring for the purpose of judging and of determining “value,” by comparison to a benchmark or standard. Evaluation refers to a program, course, or institution, whereas assessment refers to an individual.³

The AO Foundation, based in Switzerland, is a medically guided, nonprofit organization that delivers more than 700 face-to-face and online educational events around the world annually, supported by nearly 14,000 volunteer medical faculty and attended by more than 55,000 participants in many clinical areas. The AO's educational activities are designed using a competency-based approach through backward planning⁴ and follow Kern's 6-step model for curriculum development. Kern recommends assessment for planning and evaluation as a driving force for continuous improvement.³

In 2003, the AO implemented a structured evaluation system focused on the relevance of educational activities and the performance of faculty instructors. Evaluation data were summarized by the course organizer and provided to the appropriate course chairperson in preparation for the following year's events. Evaluations were limited to a small proportion of the courses (mostly international), and the data from these evaluations were not fully used. In addition, every AO clinical division (CD) and region was using different evaluation methods and tools. The uncoordinated application of evaluations was possibly the result of the insufficient tailoring of questions to the specific needs of the user.⁵ Faced with this inconsistent data collection, planning committees were unable to properly apply the program evaluation findings to make decisions regarding course planning and improving the curricula. At the same time, CME providers were increasingly pressured by leaders in medical education, governments, health authorities, and accrediting bodies to provide evidence of educational outcomes.^{3,6–9} Providers were asked to perform and document needs analyses (gap analyses), to define learning objectives, demonstrate independence of education, show ongoing improvement, and collect data on the changes (outcomes) that result from educational interventions. It was therefore essential that providers have access to structured and consistent collection and reporting of

Accepted for publication November 10, 2020.

From the AO Education Institute, AO Foundation, Duebendorf, Switzerland. This project was supported entirely by AO internal funding.

The authors have no conflicts to report. The authors serve as volunteer surgeon faculty at AO North America and have no financial and/or personal interests in this work.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jorthotrauma.com).

Reprints: Monica Ghidinelli, PhD, AO Foundation Education Institute, Stettbachstrasse 6, 8600 Dübendorf, Switzerland (e-mail: monica.ghidinelli@aofoundation.org).

Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

evaluation and assessment data to enable accurate analysis and to optimize planning decisions.

For these reasons, the AO decided in 2012 to update and align all the questions and processes for gathering and reporting data before and after educational events. The AO instituted the new system in 2016. This article describes the process that was used to design, develop, and implement a more effective evaluation and assessment system for educational events worldwide.

METHODS

Based on the intended use of the course data, an Evaluation for Quality Improvement, characterized by intensive stakeholder engagement,¹⁰ was the best suited approach. A series of expert advisory group sessions and consensus meetings were convened over a 3-year period. Methods included literature review,^{4,11–13} evidence gathering, consensus-building debates, and meetings with stakeholders to ensure that contributions reflected the various perspectives and that all aspects were relevant to end users. The process consisted of the following 5 phases:

Phase 1: Planning

1. A steering team was formed from a group of experts in medical education and evaluation and a plan of action (goals, contributions, timeline) was established.
2. Through an interview process, stakeholders were identified and asked to define the factors that contribute to the success of educational activities and the level of outcomes⁴ they wanted to assess. A list was collated, and the highest scored factors were combined with existing educational program evaluation frameworks^{4,11–13} to create a proposed set of data collection areas.
3. The proposed stakeholder groups and data collection areas were confirmed at the AO Education Platform annual meeting where all AO clinical divisions (CD) are represented.
4. The CD representatives shared the proposal with the respective CD, and feedback was integrated in the proposal.

Phase 2: Question Development

1. The steering team compiled a list of questions for each of the data collection areas, combining previously used questions, input from stakeholders, and evidence from the literature.^{11–14} The guiding principles for each question were that they must be clearly understood by an international audience, provide a sociodemographic profile of the responders, provide quantifiable and actionable information, and be answered in a short amount of time.
2. The list was reviewed by members of the CD and feedback was integrated (see questions and descriptions in **Appendices 1–3, Supplemental Digital Content 1**, <http://links.lww.com/JOT/B301>, <http://links.lww.com/JOT/B302>, <http://links.lww.com/JOT/B303>).

Phase 3: Report Design and Action Collection

1. A proposal for report content, format, and timing was prepared by the steering team.
2. The proposal was reviewed and consolidated during the AO Education Platform annual meeting according to priority and resource availability.
3. Reports were piloted at the AO Davos Courses in 2013 and feedback from faculty and chairpersons regarding quality, thoroughness, and potential actions were collected and integrated (see sample reports and descriptions in **Appendices 1–3, Supplemental Digital Content 2**, <http://links.lww.com/JOT/B301>, <http://links.lww.com/JOT/B302>, <http://links.lww.com/JOT/B303>).

Phase 4: Development of Online Tools and Workflow

1. A proposal for tools and workflow was created by the steering team and eLearning/IT experts.
2. The proposal was discussed with the CD, and agreement was reached about content, timing, and roles.
3. IT interfaces and automation for the process were developed.

Phase 5: Implementation

1. A rollout plan was developed by the steering team.
2. In 2015, the pilot was tested at AO Dubai's regional courses and AO Davos' international courses, followed by final adjustments.
3. A communication plan was developed and outreach to the AO community began.
4. Training was organized for data collection staff and report recipients.
5. Support was established and an online guide was created.
6. Translations of questions and reports were prioritized based on the need.
7. A cost–benefit analysis was made of all the required changes to the existing evaluation system.

Data Collection and Management

Data collection, handling, and management were conducted according to General Data Protection Regulation (GDPR) standards. Participant data were anonymized and available only in aggregate form. The applications used were SurveyMonkey (to collect data via questionnaires) and Tableau (to analyze and visualize data). A custom-built and automated system controlled the time-sensitive workflows for the distribution of questionnaires and reminders, and the management of collected data for reporting.

RESULTS

The overarching goals for the creation of an evaluation and assessment system for the AO were to (1) measure the impact of educational activities on the competence of the surgeon learners and the faculty, (2) measure the effectiveness of planning decisions and achievement of learning outcomes, and (3) inform future iterations of educational activity planning and meet new or updated needs.

The first step was to identify and engage all the groups that might be interested in the results. Eight

stakeholder groups were identified: participants, faculty, chairperson(s) (course chairperson, coaches, educators, supervisors), curriculum developers (faculty, taskforces, educators, staff), management, boards/councils (involved in strategic decisions for educational activities), funding bodies and partners, and CME accrediting authorities. The stakeholders then defined the factors for successful educational activities: (1) participants demonstrate improved knowledge, skills, or attitudes and a change in practice or behavior, (2) faculties are well prepared and effective, and (3) content is relevant and commercially unbiased. These success factors represented the basic principles that were used in combination with the literature to specify the 7 data collection areas and the 19 questions (7 pre-event and 12 post-event questions) that constitute the standard data set. These are used in all regions and surgical specialties with available adaptation to clinical areas and serve the minimal reporting requirements. The standard questionnaire can be expanded with 5 optional set of questions.

Data Collection Areas

Area 1: Demographics

Demographics data provide information about the level of experience and expertise of participants and their backgrounds (specialties or subspecialties and type of practice) (see **Appendix 1, Supplemental Digital Content 3**, <http://links.lww.com/JOT/B301>). This is relevant for chairpersons and faculty before the educational event and for all the stakeholders to effectively compare course results with each other.

Area 2: Motivation

Motivation is the fundamental precondition for successful learning and is based on needs.¹⁵ Motivational data are collected before and after the event to measure possible changes in needs of certain competencies.¹⁴ To estimate motivation, the gap (difference) between self-reported desired level of expertise and current level of expertise is calculated for each defined event competency (or learning objective).^{14,15} An optional addition is an objective measure using 2 multiple-choice questions (MCQs) for each competency. This enables detection of areas where actual needs differ from the perceived ones¹⁴ (see **Appendix 1, Supplemental Digital Content 4**, <http://links.lww.com/JOT/B301>). Knowledge of motivation is essential for faculty before the educational event, chairpersons and management for future event planning, curriculum developers when making adjustments, and CME accreditation authorities.

Area 3: Faculty Performance

Faculty performance must be good for the successful rating of an educational activity. The standard data set can be extended to faculty performance for each lecture, discussion, and practical exercises (collected during the event with paper and pencil) (see **Appendix 2, Supplemental Digital Content 5**, <http://links.lww.com/JOT/B302>). This information can be applied by the faculty to assess their own performance and to help chairpersons with future faculty selection for similar educational activities.

Area 4: Event Key Performance Indicators

Event key performance indicators (KPIs) like venue/location, communication, and perceived commercial bias of the event are important factors to consider after the event by management (especially course organizers), chairpersons, and CME accreditation authorities. These factors may influence participants' recommendation of the course (see **Appendix 2, Supplemental Digital Content 6**, <http://links.lww.com/JOT/B302>).

Area 5: Outcome Participation

Outcome participation represents the number of participants progressing through each stage of the participation funnel, for example, how many registered, showed up at the event, attended all the sessions, and completed evaluations.^{4,16} This information is essential after the event for management (especially course organizers) and curriculum developers. These data are collected without asking direct questions of participants (see **Appendix 2, Supplemental Digital Content 7**, <http://links.lww.com/JOT/B302>).

Area 6: Outcome satisfaction

Satisfaction measures the degree to which the expectations of the learners about the educational activity were met.⁴ These measures are used by chairpersons, curriculum developers, management (especially course organizers), and CME accreditation authorities. An option is available to expand the data set and rate the relevance of the content for each lecture, discussion, and practical exercises (see **Appendix 2, Supplemental Digital Content 8**, <http://links.lww.com/JOT/B302>).

Area 7: Outcome Learning, Competence, and Performance

Achieving an increase in competence and performance is the gold standard in today's medical education.⁴ This information is relevant for all the stakeholders, especially chairpersons, management, curriculum developers, and CME accreditation authorities.

To estimate learning, participants are asked to self-report knowledge gain. An optional objective measure is available of a set of 2 MCQs for each competency and are comparable to the pre-event questions (see **Appendix 2, Supplemental Digital Content 9**, <http://links.lww.com/JOT/B302>).

To estimate an increase in competence, participants are asked to describe 1 to 3 specific changes they intend to make in their clinical practice and relate them to competencies. This provides an opportunity for self-reflection, which in itself promotes learning.¹⁷ In addition, self-reported current level of ability for each competency was compared with the one provided before the activity.¹⁸ As an expanded option, to estimate change in performance, a Commitment to Change follow-up questionnaire 3 months after the event is available.¹⁹ Participants are also asked to self-report the implementation status of the intended changes and barriers (see **Appendixes 2 and 3**,

Supplemental Digital Content 10, <http://links.lww.com/JOT/B302>, <http://links.lww.com/JOT/B303> respectively).

The layers of data collection allow for adaptation to different needs and offer different options for assessment, which increase reliability and outcome levels (Table 1).

The 19 standard and 5 optional questions were used to generate 2 standard and 4 optional reports aimed at different recipients (Table 2) (see **Appendices 1–3, Supplemental Digital Content 11**, <http://links.lww.com/JOT/B301>, <http://links.lww.com/JOT/B302>, <http://links.lww.com/JOT/B303>). Additional reports based on needs can be generated on demand by aggregation of different data sets.

IT infrastructure and a highly automated workflow were developed and piloted in 2015 for the entire evaluation and assessment process to collect, combine, and analyze data for one event (Fig. 1) or several events over time. An option for the collection of data with paper was also provided to respond to regional needs and to increase the response rate. The standard set of question and reports were initially provided only in English and then translated into Spanish, Portuguese, Chinese, Russian, French, German, Italian, Japanese, and Korean.

In 2016, chairpersons, faculty, and staff in each region were informed about the new system, and course organizers and report recipients were trained on how to apply the information. In particular, in the chairperson education program (CEP), chairpersons were trained on the use of pre-event participant data reports to adjust course content to participant levels and to share the data with faculty during the precourse meeting (see **Appendix 4, Supplemental Digital Content 12**, <http://links.lww.com/JOT/B304>). They also learned how to analyze all the other report types.

In mid 2017, all CD and regions were asked to fully implement the pre- and postevent reports as standard for all face-to-face events and to adapt them for online activities (webinars, webcasts, etc.). The other report types remained optional depending on local needs and resources.

In October 2018, the new evaluation system was mandated for all courses and seminars, and in 2019, it was implemented in more than 70% of AO Trauma educational events globally.

DISCUSSION

The AO created a highly automated workflow for structured and consistent collection and reporting of evaluation and assessment data. The mix of multiple-choice and open-field questions provides quantitative and qualitative data that can be used for course improvement and research studies.

This agile process, together with the engagement of stakeholders during development, helped to ensure the commitment in using the reports and in taking action on findings.²⁰ The event reports are used by chairpersons and faculty to adjust individual events and meeting the needs of participants. All groups can use these reports to monitor the overall success of the educational activity for improved competence and performance. In addition, course organizers can easily meet the standard for reporting to the CME accrediting authorities.

When monitoring and planning educational events, most stakeholders compare reports over time or with similar events. For example, curriculum developers regularly use aggregated data to monitor new curriculum performance and implementation.¹⁸ This is of particular interest to global providers because the AO faces the additional challenges of different health care systems and sociocultural environments.²¹ Furthermore, curriculum developers use the data to identify trends, new or changed participant needs, and to adapt or develop new curricula.¹⁹ An emerging trend of courses based on curriculum development is that they are rated higher in content usefulness and participant satisfaction than those that do not. Boards, councils, and management use the aggregated data for planning purposes focusing on the impact of education by course type, by region, or over time.

Recently, the increased demand for online education to address the restrictions in conducting face-to-face events because of the Covid-19 virus pandemic highlights the value of having baseline data. The use of these data enables any organization to effectively evaluate adapted course delivery through online or blended educational methods compared with their existing standards.

Challenges in the implementation of the new evaluation system were predominantly faced during the early stage. The main obstacle was language, with questions and reports first

TABLE 1. Assessment Options Available

	Pre-event	Post-event	Outcome Level
Standard	Self-reported level of ability of each competency (or objective)	Self-reported level of ability of each competency (or objective) + Commitment to change	Self-reported Learning and competence
Additional options			
1	Set of 2 MCQs for each competency (or objective)		
2		Set of 2 MCQs for each competency (or objective)	
3	Set of 2 MCQs for each competency (or objective)	Set of 2 MCQs for each competency (or objective)	Learning and competence
4		Three mo commitment to change follow-up	Performance

TABLE 2. Report Types

	Report type	Release Time	Recipients
Standard	Preevent participant data report	20, 10, 3 days before	Chairperson(s), curriculum developers, management
Standard	Postevent evaluation report	16, 30 days after	Chairperson(s), CME accrediting authorities, curriculum developers, management
Optional	Content and faculty report (confidential)	16 days after	Chairperson(s)
Optional	Individual faculty reports	16 days after	Individual faculty members
Optional	CME report	16 days after	CME accreditation body
Optional	Commitment to change outcome report	105 days after	Chairperson(s), curriculum developers

administered only in English. Translations increased usage of the evaluation system. Interpretation of the data was not always straightforward because benchmarks were not provided, and each group needed to set them by reviewing data over time and with comparisons of similar events. This was required because interpretation must always consider the language of the questions compared with the fluency of the

responders, the culture within the country or subspecialty, and the overall context of the educational event. An additional challenge that is still partially present is that content and faculty ratings by lecture are collected on paper during the event and are then manually reported on the database. Although this ensures a high response rate, it requires the event organizer’s added time and resources.

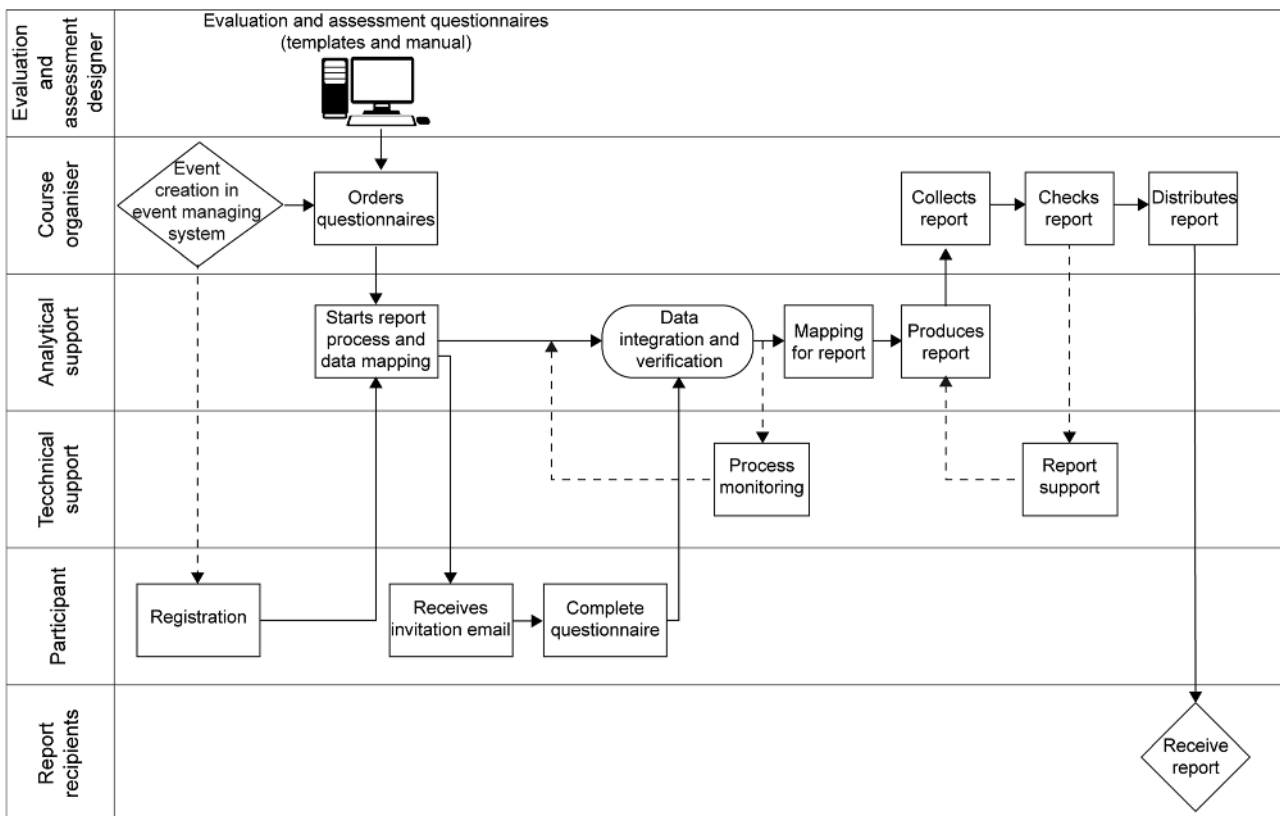


FIGURE 1. The course chairpersons request an online evaluation. The course organizers select the appropriate reports in the event managing system. Registered participants receive invitations and reminders with the SurveyMonkey link for the pre-event questionnaire by email on predefined dates. Data from registration and SurveyMonkey are stored in a database and processed. The course organizer receives and distributes the pre-event reports. During the event, faculty performance and content usefulness for each lecture, discussion, and practical exercise are collected on paper (reported manually in the database). After the event, the participants receive the pre-event questionnaire by email (SurveyMonkey link). The course organizers receive and distribute the pre-event reports. Technical and analytical support are constantly provided.

The limitations of this new evaluation system include a lack of completion by participants in some events, especially in regards to the 3-month follow-up reports, single source feedback, and possible inflexibility because of selected technical solutions and software in an area that is constantly changing, and the time and costs required for any changes. In addition, our data collection strategy is prone to voluntary response bias (eg, people with strong opinions are more likely to respond to a poll) or nonresponse bias.²²

Future enhancement of this evaluation and assessment system would be to automate the generation of yearly reports, to consider integrating alternative assessment techniques to MCQs and commitment to change (eg, case reviews, script concordance tests), and to integrate a more reliable measurement instrument for faculty performance.

ACKNOWLEDGMENTS

The authors specially thank the AO Education Platform members during project (Chairpersons: Suthorn Bavonratanavech, Nikolaus Renner, Robert McGuire. AO Trauma: Jaime Quintero, Kodi Kojima, Wa'el Taha; AO Spine: German Ochoa, Mike Grevitt, Bryan Ashman; AO CMF: Warren Schubert, AO VET: Bruno Peirone, Rico Vannini, Alessandro Piras) and the AO Education managers. AO North America gratefully acknowledges support for its education activities from the AO Foundation. The AO Foundation receives funding for education from Synthes GmbH.

REFERENCES

- Campbell C, Silver I, Sherbino J, et al. Competency-based continuing professional development. *Med Teach*. 2010;32:657–662.
- Alkin M, Taut S. Unbundling evaluation use. *Stud Educ Eval*. 2003;29:1–12.
- Kern D, Thomas P, Hughes M, et al. *Curriculum Development for Medical Education: A Six-step Approach*. Baltimore, MD: The John's Hopkins University Press; 2009.
- Moore DE Jr, Green JS, Gallis HA. Achieving desired results and improved outcomes: integrating planning and assessment throughout learning activities. *J Contin Educ Health Prof*. 2009;29:1–15.
- House E. Assumptions underlying evaluation models. *Educ Res*. 1978;7:4–12.
- Guenova M, Schafer R, Palange P. Independent continuing medical education (CME)/continuing professional development (CPD) must deliver unbiased information. *J Eur CME*. 2019;8:1690321.
- WHO. *Regional Guidelines for Continuing Medical Education (CME)/Continuing Professional Development (CPD) Activities*. WHO Regional Office for South-East Asia, 2010. Available at: <https://apps.who.int/iris/handle/10665/205767>. Accessed May 2, 2020.
- Medicine Io. *Redesigning Continuing Education in the Health Professions*. Washington, DC: The National Academies Press; 2010.
- CEPD (continuing education & professional development) strategic planning retreat participants. *Strategic Plan 2011–2016 Better Education Better Outcomes*. 2011. Available at: <https://documents.cpdutoronto.ca/>. Accessed May 3, 2020.
- Patton M. *Utilization-focused Evaluation: The New Century Text*. Thousand Oaks, CA: Sage Publications, Inc; 1997.
- Waagen A. *Essentials for Evaluation*. Info-line Alexandria, VA: ASTD; 1997:259705.
- Patton MQ. *Essentials of Utilization-Focused Evaluation*. Thousand Oaks, CA: Sage Publications; 2011.
- Colton D, Covert RW. *Designing and Constructing Instruments for Social Research and Evaluation*. Hoboken, NJ: Jossey-Bass, A Wiley Imprint; 2007.
- de Boer PG, Buckley R, Schmidt P, et al. Learning assessment toolkit. *J Bone Joint Surg Am*. 2010;92:1325–1329.
- Fox R, Miner C. Motivation and the facilitation of change, learning and participation in educational programs for health professionals. *J Cont Educ Health Prof*. 1999;19:132–141.
- McGowan BS, Mandarakas A, McGuinness S, et al. Outcomes standardisation project (OSP) for continuing medical education (CE/CME) professionals: background, methods, and initial terms and Definitions. *J Eur CME*. 2020;9:1717187.
- National, Research, Council. *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC: The National Academies Press, 2000.
- O'Malley NT, Cunningham M, Leung F, et al. Early experience in implementation of a learning assessment toolkit in the AOTrauma geriatric fracture course. *Geriatr Orthop Surg Rehabil*. 2011;2:163–171.
- Cunningham M, Kates S, Blauth M. Using a commitment to change tool for evaluation and planning of a global competency-based curriculum in orthogeriatrics. *J Contin Educ Health Prof*. 2014;34:123–130.
- Brandon P, Fukunaga L. The state of the empirical research literature on stakeholder involvement in program evaluation. *Am J Eval*. 2014;35:26–44.
- Giuliani M, Frambach J, Driessen E, Martimianakis MAT. Exploring globalization in the construction and implementation of global curricula. *J Cancer Educ*. 2020. doi: 10.1007/s13187-020-01705-5.
- Phillips AW, Reddy S, Durning SJ. Improving response rates and evaluating nonresponse bias in surveys: AMEE Guide No. 102. *Med Teach*. 2016;38:217–228.