

Evaluating Your Evaluations: A Practical Guide

John A. Owen, EdD, MS, Ronald T. Murray, EdD, CBiol, MBIol, PGCE, Jann T. Balmer, PhD, RN, Colleen A. Conner, Judith E. Marymor, Tammy S. Rogers

Office of Continuing Medical Education, University of Virginia School of Medicine, Charlottesville, Virginia, USA

ABSTRACT

Background: Measuring the effectiveness of continuing medical education (CME) activities has increasingly necessitated relating it to improvements in health care delivery and ultimately to patient outcomes. Recent increased emphasis on patient outcomes involves moving from a traditional evaluation model to an outcomes-based evaluation model. Making this shift to an outcomes-based model requires that CME providers develop a process for “evaluating their evaluations.”

Methods: We designed and completed a qualitative study of our evaluation instruments. Content analysis was used as the textual-analysis method for identifying coherent and important themes in the data, with the evaluation item being our unit of analysis. Once this analysis was completed, we identified the evaluation level and the updated Accreditation Council for Continuing Medical Education (ACCME) criteria that related to each evaluation item. We then developed strategies to increase the percentage of our outcomes-based evaluation items.

Results: The majority (83%) of our evaluation items pertained to learners’ participation and satisfaction, and the remaining items (17%) pertained to learners’ enhanced knowledge and skills. All of our evaluation items were categorized as traditional evaluation items. The development of strategies to increase the percentage of outcomes-based evaluation items facilitated the shift from a traditional to an outcomes-based evaluation model.

Conclusions: Completing this process of evaluating our evaluations provided us with the percentages of our evaluation items that were associated with the traditional evaluation model. The results of this analysis enabled us to develop and incorporate new strategies designed to shift us toward an outcomes-based evaluation model as a means of linking our CME activities to significant improvements in patient care.

Education never ends, Watson. It is a series of lessons with the greatest for last.

Arthur Conan Doyle, *The Adventure of the Red Circle* (1911)

CE Meas. 2008;2:74-80. doi:10.1532/CEM08.08108 | ©2008 Carden Jennings Publishing Co., Ltd.

INTRODUCTION

The evolution of expectations for continuing medical education (CME) reflects a desire to connect the education of physicians and other health care professionals to improvements in health care delivery and ultimately to patient outcomes. This transition requires a new framework not only for educational design but also for measuring the impact of these educational activities in the lifelong learning and practice improvement of physicians and health care professionals as they move through their professional lives.

Physicians’ and other health care professionals’ lifelong learning and their application of new clinical findings improve patient care and population health by reducing the gap between current medical evidence and current clinical practice [1]. Studies have revealed that CME activities are effective in closing this gap between evidence and practice [2]. CME activities that support lifelong learning require significant resources. In the United States, for example, the Accreditation Council for Continuing Medical Education (ACCME) reported that accredited providers’ expenditures for CME exceeded \$1.8 billion in 2006 [3]. Given the expected outcomes and significant resources related to these activities, it is imperative that CME providers evaluate the effectiveness of their activities, for several reasons:

- To provide accountability for effectiveness, impact, and use of resources in achieving educational goals and outcomes;

- To recognize the updated ACCME criteria related to performance and outcomes measurement (C1, C2, C3, C11, C18);
- To demonstrate worth in response to increasing external examination;
- To fulfill the need for continuous improvement; and
- To provide direction for future CME activities.

Measuring the impact and effectiveness of CME activities occurs at distinct levels, such as those found in the widely accepted conceptual scheme of Kirkpatrick and Philips for measuring learning outcomes [4-6]. This scheme involves 5 levels of evaluation for CME activities: level 1, reaction (learners’ participation and satisfaction); level 2, learning (enhanced knowledge and skills); level 3, behavior (changes in clinical behavior); level 4, results (improved performance that produces changes in patient outcomes); and level 5, changes in population health outcomes (Figure 1). Historically, CME evaluators have used the traditional evaluation model that focuses on levels 1 and 2 [7]. The updated ACCME criteria require a shift toward an outcomes-based evaluation model that focuses on levels 3 to 5 by measuring changes in learners’ actual behavior and the resulting improvements in patient and population health status.

Making this shift from the traditional evaluation model to an outcomes-based model requires that CME providers develop a process for “evaluating their evaluations.” Our purpose is to describe a process we

developed and implemented at the University of Virginia Office of Continuing Medical Education that was predicated on 5 distinct process questions:

1. What are we currently evaluating and measuring relative to the 5 levels of evaluation?
2. Of the 5 levels of evaluation, what levels of outcome could we be evaluating or measuring?
3. Based on the updated ACCME criteria, what could we be evaluating or measuring?
4. What are the gaps between what we are and could be evaluating or measuring?
5. How can we close these gaps and improve our evaluations?

To answer the first question, we designed and completed a qualitative study of our evaluation instruments. Once we obtained this baseline, we addressed questions 2 through 5 to shift the focus more to an outcomes-based evaluation model.

METHODS

We conducted a survey of all the evaluation methods being used in our office to provide us with the following information: (1) the specific evaluation instruments being used, (2) the content of each evaluation, and (3) the way each evaluation was being used (Figure 2). After completing this survey, we collected all of the evaluation instruments that were being used to evaluate each of our educational activities and compiled a comprehensive list of all the evaluation items within these instruments.

Content analysis was used as the textual-analysis method for identifying coherent and important themes in the data [8]. This method enabled us to compress many evaluation items into a smaller number of content categories for drawing inferences from our data. The evaluation item was our unit of analysis. Each item was

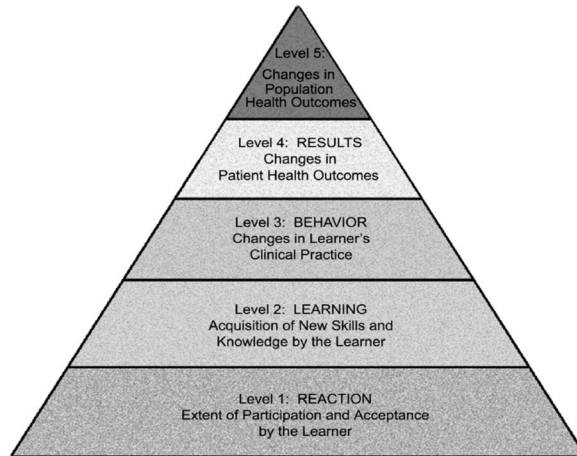


Figure 1. Five levels of evaluation for continuing medical education activities according to the scheme of Kirkpatrick and Philips [4-6].

listed on a single sheet of paper to facilitate our analysis of these items. We categorized each item according to the following procedure:

1. Coding of each item based on the 5 levels of learning outcome;
2. Naming of categories by reviewing the items and making a list of the main themes;
3. Placing items in categories based on the closest match between item content and category name;
4. Resolving as a group any disagreements in coding of items, naming of categories, and placing of items within categories;
5. Determining as a group that the categories were comprehensive (ie, covered all possibilities) and mutually exclusive (ie, nonoverlapping).

After completion of this procedure, all similarly worded evaluation items had been placed into distinct categories

Table 1. Example of Categorization Process of Level 1 Evaluation Items

Category: Educational Content	Original Educational-Content Evaluation Items	Reworded Item
Subcategory: Evidence based	Were these presentations evidence based?	The presentation contained content that was evidence based.
	The content in these presentations is most often based on evidence.	
	The presentation was evidence based.	
	Is this presentation evidence based?	
	The content in this presentation was evidence based.	
Subcategory: Material	Proportion of new material.	The presentation included new material.
	Has this material been previously covered elsewhere? If so, where?	
	The presentation contained new material.	
Subcategory: Quality	Quality of Information?	Overall, the quality of the presentation was excellent.
	The quality of the presentation overall was excellent.	

1. Of the following list of education activities, please indicate those activities in which you are involved:

- a. ☐ Grand Rounds
- b. ☐ Clinical Connections
- c. ☐ Online Activities
- d. ☐ Live Conferences
- e. ☐ Affiliates and Outreach Program
- f. ☐ Enduring Materials

2. Please list any additional education activities in which you are involved:

3. Do you use the standard evaluation form for your education activity?

- a. ☐ Yes
- b. ☐ No
- c. ☐ I also use other evaluation methods and/or instruments.

If box "c" is checked, please either provide a brief description or attach an example of one or more of the ways you evaluate education activities.

4. Are you using other types of evaluations other than education-activity evaluations, e.g., evaluation of course directors, evaluation of speakers, and so on?

- a. ☐ Yes
- b. ☐ No

If box "a" is checked, please either provide a brief description or attach an example of one or more of these other types of evaluations.

Figure 2. Evaluation survey questions.

for each level (eg, Level 1: Educational Content [category name]). We then named subcategories by reviewing the items within each category and made a list of the main themes. We placed similarly worded evaluation items into these subcategories (eg, Educational Content [category name]: Evidenced-based [subcategory name]). This subcategory grouping of evaluation items allowed us to identify redundant items; such items were reworded to minimize redundancy. Table 1 provides an example of this categorization process.

Next, we grouped all of our CME activities into 3 major categories: Single Live, Enduring, and Regularly Scheduled. For each reworded evaluation item, we

determined which items would be required, optional, or nonapplicable for each of these 3 major categories. Evaluation items that were required for all 3 categories became our core set of evaluation items and were placed on the standard evaluation form. This step ensured that data collection and evaluation would be uniform across the entire CME program. Evaluation items classified as "required" for one of the 3 major categories of CME activity, such as Single Live, were added to this core set of evaluation items. The core set of evaluation items plus the required items that pertained to one of the 3 major categories formed the evaluation instrument specific for that category of CME activity (Table 2).

Table 2. Required Core Set of Evaluation Items and Required Evaluation Items by Activity*

Evaluation Items	Live	Regularly Scheduled	Enduring
Overall, the content of this activity was valuable to me.	Required Core	Required Core	Required Core
The educational objectives for this activity were met.	Required Core	Required Core	Required Core
This information will influence my practice of medicine.	Required Core	Required Core	Required Core
Please indicate any changes you plan to make in your practice of medicine as a result of information you received from this activity.	Required Core	Required Core	Required Core
How do you think these changes will affect patient outcomes?	Required Core	Required Core	Required Core
Please indicate if this activity was free from commercial bias (Y = yes, N = no). If you checked "no," please indicate the topic(s) that were not free from commercial bias:	Required Core	Required Core	Required Core
Please provide general comments regarding this activity and suggest how it might be improved.	Required Core	Required Core	Required Core
Disclosure of any relevant financial relationships of the presenter was made prior to this activity (Y = yes, N = no).	Required Core	Required Core	Required Core
Speakers were effective.	Required	Required	N/A
Discussion was valuable.	Required	Required	N/A
If applicable, the handouts were effective.	Required	Required	N/A
If applicable, audiovisual aids were effective.	Required	Required	N/A

*N/A indicates not applicable.

In addition, we developed a list of optional evaluation items that could be added to a required set of evaluation items (Table 3).

Our final step was to identify the evaluation level and the updated ACCME criteria that related to each reworded evaluation item. This step enabled us to determine the percentage of evaluation items that pertained to each level of evaluation and to each updated ACCME criterion. We then addressed process questions 2 through 5 to develop strategies to increase the percentage of evaluation items in levels 3 through 5 and correspondingly decrease the percentage of evaluation items in levels 1 and 2.

RESULTS

We identified 17 evaluation instruments and 130 evaluation items that were being used to evaluate our educational activities. Of the 130 evaluation items, we categorized 108 (83%) as measuring level 1 outcomes, 22 (17%) as measuring level 2 outcomes, and 0 items as measuring outcomes of levels 3 through 5. After identifying and rewording redundant evaluation items, we reduced the total from 130 to 27 distinctly different items. Of these 27 items, we categorized 15 (56%) as measuring level 1 outcomes, 12 (44%) as measuring level 2 outcomes, and 0 items as measuring outcome levels 3 through 5.

Of the 27 items, we identified 8 (30%) as being required for all 3 categories of educational activities based on the ACCME accreditation criteria, and these 8

items became our core set of evaluation items. Of these 8 items, we categorized 5 (63%) as measuring level 1 outcomes, 3 (37%) as measuring level 2 outcomes, and 0 items as measuring outcomes of levels 3 through 5. These 8 items met various relevant updated ACCME criteria, with the exceptions of criteria 6 and 18 (Table 4).

After addressing process questions 2 through 5, we determined that we could decrease our percentage of level 1 outcomes and increase our percentage of level 2 and level 3 outcomes. We desired to measure levels 4 and 5 but recognized that assessing these 2 levels of outcomes would be extremely challenging and would require long-term planning.

DISCUSSION

CME offices and organizations vary greatly and for this reason are capable of measuring different types and levels of outcomes. The University of Virginia Office of Continuing Medical Education, as part of the Dean's Office, serves as a strategic and organizational resource whose activities are designed to foster the development of skills and knowledge in medicine and related fields, promote standards of care, and facilitate the translation of medical and scientific research into the delivery of quality patient care for physicians and health care professionals. These CME-sponsored activities include departmental conferences, regularly scheduled subspecialty conferences, grand rounds, conferences jointly sponsored with professional health care organizations, international

Table 3. Optional Evaluation Items

The educational objectives for this activity were clearly stated.
What did you learn? How will you apply this in your practice of medicine?
The content in this activity was relevant to my practice of medicine.
The format of this activity was conducive to my learning.
Please identify specific areas of your practice that you find most challenging.
The information presented reinforced my current practice/treatment habits.
As a result of attending this CME activity, I declare: • My general medical knowledge was enhanced. • This education is within my scope of practice; however, I'm uncertain, at this time, if I will make a change in my performance. • Or, I will be making the following changes in my performance: (Develop check list from educational content and objectives).
The information presented in this educational activity enhanced my current knowledge base.
The information presented in this educational activity addressed my most pressing questions.
The information presented in this educational activity provided new ideas or information I expect to use.
The information presented in this educational activity addressed competencies identified by my specialty.
Please list any topics you would like to see addressed in future educational activities:
Name one thing you intend to change in your practice as a result of completing this educational activity.
Do you expect that the information you learned during this activity will help you improve your skills or judgment within the next 6 months? (1 [definitely will not change] to 5 [definitely will change]).
How confident would you feel treating patients in your own practice based on the activity content (1 [not very confident] to 5 [very confident/certain]).

and national meetings, affiliated hospital activities, and enduring materials, such as print and computer-assisted instructional materials. Many of our activities are large in scope. In our 2007 ACCME Annual Report, for example, we list 94 directly and jointly sponsored live courses (1428 hours of instruction, 6856 participants) and 99 directly and jointly sponsored regularly scheduled series (1880 hours of instruction, 34,464 participants).

Through our analysis, we determined that the highest percentage of our current evaluation items measured level 1 outcomes and that none of our current evaluation items measured levels 3 through 5. This result revealed that we were using a traditional evaluation model. Historically, we have included 2 evaluation items to assess what participants learned and what changes they planned to make in their practices. These planned changes in practice behavior are known as commitment-to-change (CTC) statements. We recognized that these 2 items assessed knowledge acquisition and should be classified as level 2 learning outcomes. Although helpful for participant self-reflection and goal setting, they do not measure self-reported changes in practice behavior; however, self-reported changes in practice behavior months following the educational activity would be classified as level 3 learning outcomes.

We identified numerous strategies that would increase the percentage of level 2 and level 3 outcomes relative to the updated ACCME criteria. To document changes in learning, we included pre/post tests, skill observation, and

CTC statements in strategies pertaining to level 2 outcomes. Strategies pertaining to level 3 outcomes included rewriting several of our evaluation items and conducting CTC follow-up studies 6 to 12 months after the activity. Through conducting these follow-ups, it would be possible to assess the true impact of the activity on changes in clinical-practice behavior and to gain an idea of unintended changes and barriers to change. We recognized, however, that the value of self-reported data may be limited and that evaluating levels 4 and 5 still would be needed.

We found that developing strategies to measure level 4 and level 5 outcomes was extremely challenging. We did note, however, that the available data provided opportunities for measuring these 2 levels. These data included measures of quality, utilization, screening, diagnoses, prescribing patterns, imaging, hospitalization, patient satisfaction, and even public health. In addition, we recognized that Performance Improvement CME and collaborating with our Health System's Quality Improvement Committee might provide additional strategies for measuring these higher outcomes. Data specifically available for measuring level 5 outcomes included measures of treatment outcomes and health status, such as morbidity and mortality, secondary complications, adequate control of underlying disease, hospitalization and rehospitalization, and community public health statistics.

Finally, we acknowledged that both lower and higher levels of learning outcomes needed to be evaluated,

because the validity of the data gathered at each level depends on the validity of the data gathered at the level below it. We also recognized that measuring these different levels provided a rich source of information that could be used in planning future CME activities. For example, information from level 1 outcomes could be used during the initial stages of instructional design. Information from level 2 outcomes, especially CTC statements, could give some idea whether participants understood key messages and were ready to implement change. Information obtained from level 3 measures could document actual practice change, and, finally, information from measures of levels 4 and 5 could indicate whether patients' health outcomes had improved.

Next Steps

Due to the updated ACCME criteria, our educational-planning process has shifted to an integrated outcomes-based learning model. This new planning process includes assessing gaps in physician performance, creating outcomes-based learning objectives, implementing methods and formats that have an impact on learning, and evaluating outcomes. Historically, our learning objectives have contained action verbs, were divided up by the content,

and were focused on the learner, and our evaluation items measured participant satisfaction and perceptions of possible changes in practice behavior. Outcomes-based learning objectives, however, are based on the desired outcomes and on application to clinical practice. Correspondingly, outcomes-evaluation items and strategies are designed to measure higher levels of learning, assess participants' perception of changes in practice behavior over time, determine changes in physician performance, and measure changes in patient and population health outcomes. This new educational-planning process has had a profound influence on our next steps, which are as follows:

1. Rewrite several of our standard evaluation items to measure higher levels of outcome and specific updated ACCME criteria (ie, evaluate outcomes in terms of increases in knowledge, improvements in attitudes, or enhancements to clinical competence and/or performance).
2. Develop a conceptual framework and methodology for course directors and CME educational planners to follow in creating outcomes-based learning objectives and evaluation strategies based on the desired outcomes identified during the needs assessment.

Table 4. Categorization of Core Set of Evaluation Items Based on the Updated ACCME Criteria and the 5 Levels of Learning Outcome*

ACCME Criteria	Level (1-5)	Required Items: Live	Required Items: Regularly Scheduled	Required Items: Enduring
4	1	Overall, the content of this activity was valuable to me.	Same item	Same item
2, 3, 5	1	The educational objectives for this activity were met.	Same item	Same item
3, 11	2	This information will influence my practice of medicine.	Same item	Same item
3, 11	2	Please indicate any changes you plan to make in your practice of medicine as a result of information you received from this activity.	Same item	Same item
3, 11	2	How do you think these changes will affect patient outcomes?	Same item	Same item
7, 9, 10	1	Please indicate if this activity was free from commercial bias (Y = yes, N = no). If you checked "no," please indicate the topic(s) that were not free from commercial bias:	Same item	Same item
2, 3, 4, 5	1	Please provide general comments regarding this activity and suggest how it might be improved.	Same item	Same item
7, 9, 10	1	Disclosure of any relevant financial relationships of the presenter was made prior to this activity (Y = yes, N = no).	Same item	Same item

*ACCME indicates Accreditation Council for Continuing Medical Education.

3. Create a database for entering, storing, and analyzing outcomes measures from our activities. The results would be aggregated, and reports could be generated (eg, a report to indicate how participants have applied what they have learned to their practice). The outcomes measures would be tied back to needs-assessment data, thereby completing the linkage process.
4. Develop strategies for assessing the changes in participants' practice several months following the activity. For example, at the end of the activity, we would ask participants to identify changes they will make in their practice of medicine by using a CTC strategy. As a follow-up, we would measure the changes in physicians' perceptions of their own behavioral changes at several different times after the activity. Another strategy would involve sending open-ended questions and/or specific item-response questions related to the changes participants have made in their practice of medicine as a result of the activity. A third strategy would involve asking participants to respond to a Case Study several months following the activity. All of these strategies would include an opportunity for participants to identify barriers they encountered in implementing their changes into their practice.
5. Develop strategies for collecting and analyzing objective measures of physician performance and patient health outcomes. For example, we could offer Performance Improvement CME to a small group of physicians who had participated in an educational activity. We also could collaborate with our Health System's Quality Improvement Committee to develop and evaluate educational interventions designed to address areas of identified need within the hospital.

CONCLUSION

Completing this process of "evaluating your evaluations" requires a commitment of time and resources; however, there are several outcomes that can be expected

as a result of completing this process. One outcome includes determining the percentage of evaluation items associated with each level of evaluation, removing redundant evaluation items, developing a required core set of evaluation items across all activities, and incorporating strategies that reduce the percentage of level 1 evaluation items and increase the percentage of level 2 and level 3 evaluation items, thereby making the shift to an outcomes-based model of evaluation. Another outcome is the completion of a CME program-improvement project that relates to the requirement of ACCME criteria 11 through 15 that accredited providers evaluate the effectiveness of their overall program and make improvements to their program. In summary, our goal in describing this process is to provide information to CME providers who are shifting toward an outcomes-based evaluation model as a means of linking their CME activities with significant improvements in patient care.

REFERENCES

1. Committee on the Health Professions Education Summit Institute of Medicine. Greiner AC, Knebel E, eds. *Health Professions Education: A Bridge to Quality*. Washington, DC: National Academy Press; 2003.
2. Marinopoulos SS, Dorman T, Ratanawongsa N. *Effectiveness of Continuing Medical Education*. Baltimore, MD: The Johns Hopkins University Evidence-Based Practice Center; 2007. AHRQ Publication No. 07-E006.
3. Accreditation Council for Continuing Medical Education Web site. Available at: <http://www.accme.org>. Accessed July 21, 2008.
4. Kirkpatrick DL. *Evaluating Training Program: The Four Levels*. 2nd ed. San Francisco, Calif: Berrett-Koehler; 1998.
5. Kirkpatrick DL. Revisiting Kirkpatrick's four-level model. *Train Dev*. 1996;50:54-57.
6. Phillips JJ. *Return on Investment in Training and Performance Improvement*. 2nd ed. Boston, Mass: Butterworth-Heinemann; 2003.
7. Havens C, Bellman P, Jayachandran K, Waters S. Measuring higher-level outcomes. *Almanac*. 2005;27:1-3.
8. Patton MQ. *How to Use Qualitative Methods in Evaluation*. 2nd ed. Newbury Park, Calif: SAGE Publications, Inc; 1987.