Evaluating technology-enhanced learning: A comprehensive framework

DAVID A. COOK1 & RACHEL H. ELLAWAY2
1Mayo Clinic College of Medicine, USA, 2Northern Ontario School of Medicine, Canada

Abstract

Background: The absence of a standard, comprehensive approach to evaluating technology-enhanced learning (TEL) limits the utility of individual evaluations, and impedes the integration and synthesis of results across studies.

Purpose: To outline a comprehensive framework for approaching TEL evaluation in medical education, and to develop instruments for measuring the perceptions of TEL learners and instructors.

Methods and results: Using both theoretical constructs of inquiry in education and a synthesis of existing models and instruments, we outlined a general model for evaluation that links utility, principles, and practices. From this we derived a framework for TEL evaluation that identifies seven data collection activities: needs analysis; documentation of processes, decisions, and final product; usability testing; observation of implementation; assessment of participant experience; assessment of learning outcomes; and evaluation of cost, reusability, and sustainability. We then used existing quality standards and approaches to develop instruments for assessing the experiences of learners and instructors using TEL.

Conclusions: No single evaluation is likely to collect all of this information, nor would any single audience likely find all information elements equally useful. However, consistent use of a common evaluation framework across different courses and institutions would avoid duplication of effort and allow cross-course comparisons.

Introduction

Although we have both previously tackled the challenges of evaluation in medical education (Ellaway 2006, 2010a; Cook 2010), we have observed ongoing challenges in the review and evaluation of technology-enhanced learning (TEL), both in practice and in the literature. These challenges include a lack of comprehensiveness or a clear focus on the function of the evaluation, unclear relationships between what happened and what was evaluated, and the absence of a sound conceptual grounding for the subject or the methods of the evaluation.

Previous authors in medical education have proposed guidelines (Atkins & O’halloran 1995; Glenn 1996; Olson & Shershneva 2004) and specific instruments (Knight et al. 2004; Alyusuf et al. 2013) for judging the quality of computer-based learning materials. Standards and benchmarks for online instruction have also been proposed for education generally (Merisotis & Phipps 2000; Oliver 2000; Moore 2005; Quality Matters Program 2014). However, none of these instruments or frameworks offers a comprehensive view of the purposes, functions, and specific activities required to fully evaluate a TEL intervention. The absence of a standard, comprehensive evaluation approach limits the utility of individual evaluations, and impedes the integration and synthesis of results across studies.

Our purpose is to outline a comprehensive framework for approaching TEL evaluation in medical education. We describe the development of this framework grounded both in theoretical constructs of inquiry in education and in a synthesis of existing practices and instruments. We further explain how it can be adapted to different contexts and uses in medical education, and illustrate its application in developing novel instruments for measuring the perceptions of learners and instructors as participants in a TEL course. Although we use the term “course” to refer to the unit of instruction being evaluated, this framework could equally apply to brief, focused TEL modules or multi-course programs.

Our primary intent is to support practicing teachers and program managers who need to evaluate TEL courses, although we also hope to support broader academic inquiry by encouraging greater standardization of TEL evaluation within the field of medical education. If investigators employ
The importance of evaluating technology-enhanced learning

TEL is a broad field that covers all uses of digital technology to support and mediate educational activities (Goodyear & Retalis 2010). TEL can involve web-based technologies, mobile devices and apps, computers, tablets, and other digital devices, and may include activities that are entirely digitally-mediated or those that integrate technology into hybrid or blended activities. TEL may be collaborative or self-directed, didactic or practice-based, and involve formative or summative assessment. Learners may work from home, a library, or a patient’s bedside. Most modern medical education activities involve at least some component of TEL, and as such the evaluation of TEL has become an essential – if underdeveloped – part of medical education practice.

The practice of program evaluation in medical education has been previously described, in particular the role of the evaluator (Goldie 2006) and the importance of using evaluation findings to inform the ongoing development of programs (Wall 2010). Yet, we see a need to focus on the evaluation of TEL in medical education for the following reasons:

1. There is often great anxiety about new models of medical education. Using new technologies tends to amplify this anxiety, especially given the investments in infrastructure and faculty development required for their successful integration. Responding to these stakeholder concerns necessitates rigorous evaluation of TEL activities.

2. Aspects of TEL such as usability, accessibility, and technical reliability of materials and learning environment are often overlooked in mainstream evaluation practice but play a much more substantial role in TEL evaluation.

3. Because interactions between teachers and learners in TEL tend to be different (i.e. mediated by technology), teacher–learner relationships do not yield the same opportunities for informal evaluation and feedback as do traditional modalities. This suggests that TEL requires a more comprehensive evaluation approach than other educational activities.

4. TEL can generate much more data (and different kinds of data) than traditional educational approaches. The development of educational analytics and “big data” analysis techniques affords new approaches to supporting evaluation, both procedurally and conceptually (Ellaway et al. 2014b).

5. Emerging discourses in TEL theory and practice distinguish between evaluating an educational technology in its own right and evaluating it in different contexts of use (Ellaway et al. 2014a). A greater precision in what exactly is being evaluated (and how) leads to different lines of inquiry. There are clear differences between evaluating TEL as a technology (is it reliable, safe, aligned to local environment, and sustainable?), evaluating its educational content (is it accurate, current, and in adherence to principles of effective learning?), and evaluating it as part of an overarching educational activity (did it achieve overall objectives and outcomes?) (Ellaway 2014).

Needs, uses, and guiding questions

The first step in planning any evaluation activity is to ask: “Who is the intended audience?”, and “What is the audience likely to do with the information the evaluation provides?” (Cook 2010). These questions clarify the evaluation’s purpose, and guide subsequent decisions about data collection, analysis, and presentation. Box 1 summarizes several of the evaluation frameworks noted below.

Uses

An evaluation might serve one audience and use, or many. Identifying specific uses and designing the evaluation to meet these from the outset will focus the evaluation, clarifying the information required and analyses to be conducted. For example, an evaluation might require different data collection, analysis, and presentation if intended for medical students (e.g. to inform their choice of school or selection of elective courses), instructors (e.g. to help them improve a course for the next offering), or a funding agency (e.g. to determine the course’s impact). Clarifying the intended use early on will help to save time and trouble later in the process.

The purposes of an evaluation can be broadly classified as summative or formative. Summative evaluation occurs at or after the end of a course, and renders a final judgment such as, “How well did it work?” Formative evaluation typically occurs
at various points before, during, and after a course, and aims to collect information to improve a course as it is running and in future iterations.

It is also helpful to consider how an educational intervention influences outcomes for, or is experienced by, stakeholders at the individual participant level, the group or class level, and the institutional, professional or societal level. At the level of the individual learner, cognitive events might be the primary consideration: what did they experience or learn? At the level of the group, observable activity and aggregate performance might be the focus (e.g. what went on amongst the learners and between learners, instructors, technologies, and other materials; and how well did they perform as a whole?). At the level of the institution or society we might ask about the context that gave rise to the intervention, and the value or impact on people and systems outside the classroom (e.g. the contribution to the mission of a school or to the goals of a national initiative).

Questions

Asking good questions to guide the process of inquiry is a core feature of good evaluation. Without a focused question it is difficult to judge whether an evaluation has accomplished its goals.

We can classify questions in terms of description ("What was done"), justification ("Did it work and at what cost"), and clarification ("How or why did it work, and how can it be improved") (Cook et al. 2008). Clarification questions can be further classified as those that seek explanation ("How or why did it work, or how was it experienced") and those that suggest experimentation ("How can it be improved"). We can also ask normative questions such as "What should have happened?" which we might label "judgment". This suggests a five-domain model defining key questions for planning an evaluation (i.e. description, justification, clarification-explanation, clarification-experimentation, and judgment).

### General principles of evaluation

Cook (2010) described three broad approaches or orientations for evaluation.

- **Objectives-oriented** approaches focus on how well *a priori* objectives were met. Course objectives and corresponding data might include enrollment and completion numbers, learning outcomes (knowledge, skills, attitudes), and net income and expenditures. The objectives-oriented approach is relatively straightforward in implementation, but tends to be poorly suited to capturing unexpected developments (good or bad) in the course. For example, we might discover that learning outcomes were poor, or that completion rates were unexpectedly high, and yet be unable to determine why this occurred based solely on objectives-related data.

- **Participant-oriented** approaches typically use qualitative, and occasionally quantitative, methods to collect data from multiple sources and inductively explore not only what happened, but why it happened and how it was experienced; this approach typically uses qualitative data and occasionally quantitative methods to collect data from multiple sources. This approach is responsive to unexpected events, but interpretations are typically subjective and context-specific.

- **Process-oriented** evaluation approach considers the entire lifecycle of a course or program, from the inception of the idea, through implementation and delivery, to summative judgments about quality and decisions about future iterations. The CIPP model represents a prototypical example of process-oriented evaluation.

### Glossary of evaluation frameworks cited in this article.

- **CIPP framework**: a model for program evaluation organized around the themes of Context (needs, assets, and opportunities that prompted the change and the desired goals), Inputs (alternative approaches, feasibility, and cost-effectiveness), Processes (activities and issues during development and implementation), and Products (short- and long-term outcomes, sustainability, and transportability to new settings) (Stufflebeam 2003).
- **Description-Justification-Clarification framework**: a framework for classifying the purpose of education research, first described by Henk Schmidt at the 2005 meeting of the Association for Medical Education in Europe, and then further developed and formalized by Cook et al. (2008). Description focuses on what was done; justification focuses on whether it worked; and clarification focuses on why or how it worked (or failed to work) and how it can be improved.
- **Evaluation Cookbook**: a "practical guide to evaluation methods for lecturers" that suggests specific evaluation "recipes" (methods for collecting, analyzing, and reporting information) for various situations in education; see www.icbl.hw.ac.uk/tfd/cookbook/.
- **Kirkpatrick framework**: an evaluation framework for training programs organized around four levels: level 1 reactions, level 2 learning, level 3 behaviors, and level 4 results (Kirkpatrick 1996).
- **Objectives-oriented evaluation**: an evaluation approach that focuses on how well *a priori* objectives were met. The objectives-oriented approach is relatively straightforward in implementation, but tends to be poorly suited to capturing unexpected developments (good or bad) in the course.
- **Participant-oriented evaluation**: an evaluation approach that inductively explores not only what happened, but why it happened and how it was experienced; this approach typically uses qualitative data and occasionally quantitative methods to collect data from multiple sources. This approach is responsive to unexpected events, but interpretations are typically subjective and context-specific.
- **Process-oriented evaluation**: an evaluation approach that considers the entire lifecycle of a course or program, from the inception of the idea, through implementation and delivery, to summative judgments about quality and decisions about future iterations. The CIPP model represents a prototypical example of process-oriented evaluation.
- **Quality Matters evaluation program**: an international organization focused on quality assurance of online learning, see www.qualitymatters.org.
- **SLOAN Consortium**: an international organization focused on quality online learning (now the Online Learning Consortium), see www.onlinelearningconsortium.org.
- **SWOT**: a planning and evaluation framework based on the strengths, weaknesses, opportunities and threats associated with a particular activity or program; the precise origins and first description of this framework remain obscure.
includes elements of both objectives-oriented and participant-oriented evaluation, and can be very comprehensive. However, this approach is also relatively resource-intensive, and must be initiated early in the sequence of course development.

The choice of approach depends in large part upon the anticipated uses and corresponding guiding questions. Combining approaches is often useful. For example, the evaluator might use test scores to determine the achievement of objectives, but also use a participant-oriented approach to identify unplanned events and understand the root cause of unexpected outcomes. The Context-Inputs-Processes-Products (CIPP) model (Stufflebeam 2003) is a common evaluation approach that emphasizes the process (and objectives) orientation, and also accommodates a participant orientation if desired. Context addresses the needs, assets, and opportunities that prompted the change and the desired goals. Inputs consider alternative approaches, feasibility, and cost-effectiveness. Processes focus on the actual development and implementation. Products focus on both short- and long-term outcomes as well as the sustainability and transportability (i.e. practicability of adoption elsewhere) of the intervention.

Many (but not all) educational evaluations explore quantitative outcomes. Such outcomes can be classified using the hierarchy developed by Kirkpatrick (1996). Level 1 outcomes (Reactions) look at participant satisfaction with and perceptions of the course experience and quality. These are often based on post-event self-report surveys. Level 2 (Learning) considers participants’ knowledge, skills, or attitudes in a test setting, such as a multiple-choice test of knowledge, or a skill exam using a standardized patient or a virtual reality simulator. Level 3 outcomes (Behaviors) measure performance in actual practice, such as test ordering patterns or direct observation of patient care. Level 4 outcomes (Results) assess the impact on the systems and organizations within which participants work, such as changes in patients’ health or the effectiveness or efficiency of healthcare systems. However, Yardley and Dornan (2012) observe that “different levels concern different beneficiaries” and omit others altogether (such as teachers), suggesting that Kirkpatrick-type outcomes alone will be inadequate for many contexts and applications.

Many course outcomes cannot be readily quantified. Non-quantitative outcomes such as narrative feedback about the participant experience (both learner and instructor), unplanned events (both favorable and unfavorable), and required deviations from the original implementation plan can be captured using qualitative research approaches. A holistic approach to evaluation will often employ both quantitative and qualitative measures.

**Principles into practice: Specific activities to collect evaluation data**

Having considered the high level organizing principles for evaluating TEL, we turn now to practical approaches to designing evaluations of TEL events. Evaluation requires information – information that can be analyzed, interpreted, and acted upon. In a practical sense, then, the evaluator might ask: “What information should be collected to help answer my guiding questions?” (i.e. description, justification, or clarification), followed by, “How can I collect this information?” To translate the core principles identified in the previous section into practical actions we mapped these high-level questions with the process-oriented CIPP model (see Table 1), and used an iterative consensus-building process between the two authors to identify specific elements of information that a TEL evaluation might explore.

After creating this comprehensive map of useful information elements, we again used an iterative process to identify specific evaluation activities that can generate the information required. This included setting out seven broad areas of evaluation activity, summarized in the sections that follow and

**Table 1. Map of evaluation information elements organized by guiding question (needs) and stage of CIPP model (principles)*.

| Guiding question* | Context | Inputs | Process | Product |
|------------------|---------|--------|---------|---------|
| Description (what was done) | Vision, goals, objectives | Alternative vision, goals, objectives considered | Development and implementation | Actual implementation |
| Justification (did it work, at what cost) | Assets and opportunities | Anticipated outcomes | Interim outcomes | Deviations from plan |
| Clarification: Explanation (why and how did it work; user experience) | Existing programs | Alternative instructional approaches considered; theories used to make decisions | Learner experience | Outcomes (Satisfaction, Learning, Behavior, Results) |
| Clarification: Experimentation (how to improve) | Needs and gaps | Alternative instructional approaches considered; evidence used to make decisions | Instructor experience | Cost |
| | | | | Sustainability, reusability |

*This table illustrates the creation of a conceptual framework (Step 1 in Figure 1) linking needs and questions to principles and approaches, resulting in the identification of specific elements of information.

*Questions of judgment (“What should have been done?”) constitute a secondary level of evaluation, and potentially draw upon several elements of the other questions leading to a synthesis of information and comparison with the ideal.
Table 2. Map of information elements to specific evaluation activities.

| Information element                        | Specific evaluation activity                                                                 | Area of evaluation activity |
|--------------------------------------------|-----------------------------------------------------------------------------------------------|-----------------------------|
| Existing programs                          | Needs analysis and environmental scan                                                           | 1. Conduct needs analysis and environmental scan |
| Needs and gaps                             | Needs analysis and environmental scan                                                           | 1. Conduct needs analysis and environmental scan |
| Assets and opportunities                   | Needs analysis and environmental scan                                                           | 1. Conduct needs analysis and environmental scan |
| Vision, goals, objectives                  | Documentation of processes and decisions product during planning stage                          | 2. Document processes, decisions, and final product |
| Anticipated outcomes                       | Documentation of processes and decisions product during planning stage                          | 2. Document processes, decisions, and final product |
| Anticipated beneficiaries                  | Documentation of processes and decisions product during planning stage                          | 2. Document processes, decisions, and final product |
| Usability tests                            | Usability testing                                                                             | 3. Test usability           |
| Alternative vision, goals, objectives      | Documentation of processes and decisions product during planning stage                          | 2. Document processes, decisions, and final product |
| Alternative instructional approaches considered; theories and evidence used to make decisions; unanswered questions | Documentation of processes and decisions product during planning stage                          | 2. Document processes, decisions, and final product |
| Anticipated costs for each alternative     | Interim assessment of outcomes and cost                                                         | 7. Estimate cost, reusability, and sustainability |
| Development and implementation             | Interim assessment of outcomes and cost                                                         | 7. Estimate cost, reusability, and sustainability |
| Iterative design/responsiveness            | Interim assessment of outcomes and cost                                                         | 7. Estimate cost, reusability, and sustainability |
| Interim outcomes                           | Assessment of participant experience (observation, interview/debrief, survey)                   | 5. Assess participant experience and satisfaction |
| Interim costs                              | Assessment of participant experience (observation, interview/debrief, survey)                   | 5. Assess participant experience and satisfaction |
| Unexpected events                          | Documentation of key events, and response, during implementation                               | 4. Document key events during implementation and final product |
| Actual implementation deviations from plan | Documentation of key events, and response, during implementation, including final product       | 4. Document key events during implementation and final product |
| Knowledge, skills, behaviors, systems-level | Learner outcomes assessment (K2, K3)                                                           | 6. Assess learning outcomes |
| Impact (outcomes K2-4)                     | Learner outcomes assessment (K2, K3)                                                           | 6. Assess learning outcomes |
| Cost                                       | Estimation of cost                                                                             | 7. Estimate cost, reusability, and sustainability |
| Sustainability                            | Estimation of preparation for maintenance (sustainability) and likelihood of repurposing (reusability) | 7. Estimate cost, reusability, and sustainability |
| Reusability                                | Estimation of preparation for maintenance (sustainability) and likelihood of repurposing (reusability) | 7. Estimate cost, reusability, and sustainability |

*aThis table illustrates the translation of general principles and approaches (desired elements of information) into specific evaluation activities and instruments (Step 2 in Figure 1). Note that the Information Elements are derived directly from Table 1.

*bNumbers in this column correspond to those used in the main text when discussing each evaluation activity.

cK1–K4 refer to levels in the Kirkpatrick evaluation framework (K1, reaction; K2, learning; K3, behaviors; K4, results).

in Table 2. We emphasize those activities that are relatively unique to TEL, or for which application to TEL requires special considerations.

1. Conduct needs analysis and environmental scan

Although many evaluations begin at the completion of an instructional program, a comprehensive evaluation (such as the CIPP model) actually begins when a new program or change to an existing program is first conceived. A needs analysis in the formative stages helps to clearly identify the need for change, determine whether educational change (as contrasted with adjustments to other components in the healthcare system) is the correct solution, and confirm that resources can be marshaled to implement this change (Cook & Dupras 2004).

The need for change can be determined by considering organizational/societal needs and capacities (e.g. specific knowledge and skills required by a school or hospital, or emerging regulatory requirements, political trends, and economic policies), occupational needs (the roles and specific capabilities required to meet new or existing organizational needs), and individual needs (the target group’s present performance, work environment, and ability to participate in training) (Kern et al. 1998; Training and Executive Development Group 2014). These needs should be contrasted with educational experiences and resources already existing or planned, including formal courses, clinical experiences, online knowledge resources, peers, and faculty; and with available technology resources (infrastructure and human expertise).

This information can be collected in a myriad of ways, including objective documentation of learner performance (e.g. formal tests of knowledge or skill,
direct observation of clinical performance, learner self-assessments, or chart audits), surveys, interviews or focus groups, and critical reviews of existing offerings. A group-level “SWOT” analysis can be helpful in identifying organizational Strengths, Weaknesses, Opportunities, and Threats.

Ultimately, a need for change can be justified by identifying a gap in the curriculum (i.e. inadequate depth or breadth in the specific clinical topic) or a gap in the instructional approach (i.e. failure of existing instructional strategies to optimally facilitate learning or reach desired learners).

2. Document processes, decisions, and final product

The development of a course typically blends formal instructional design models (such as 4CID [van Merriënboer et al. 2002], conversational exchange [Laurillard 2001], or significant learning [Fink 2003]) and instructional design processes (such as the Analyze, Design, Develop, Implement, and Evaluate (ADDIE) framework [Morrison et al. 2010]) with an organic response to local needs and culture. A comprehensive evaluation will carefully document each activity and decision in this process. This typically takes the form of a narrative record of the steps taken (drawing on sources such as diaries or meeting notes), the people involved and other resources required, and the nature of the resulting course. The real cost of developing and implementing TEL courses is rarely captured, but should be; see further discussion of this point below.

Specific processes or milestones that might be documented include the development of specific goals and objectives, the selection of specific instructional approaches, and the selection and configuration of specific technologies. It is also useful to document viable alternatives (e.g. objectives and instructional approaches not incorporated into the final model), and the reasons (e.g. theory, evidence, costs) and means for their exclusion.

3. Test usability

Usability focuses on the educational qualities of all the resources and tools used in a course – those things that Engeström (1987) calls “mediating artifacts” in an educational activity. Usability is likely important in all educational activities, but it is particularly relevant to TEL given its inherent dependence on technological mediating artifacts. Usability evaluation considers both the ease with which the user can perform desired activities when the technology functions as intended, and also identifies errors in function or content (Krug 2000; Nielsen 2012). Nielsen identified five key elements of usability (2012): Learnability refers to how easily users can accomplish basic tasks the first time they encounter the technology. Efficiency looks at how well users can perform tasks once they have learned the steps. Memorability refers to how easily learners can reestablish proficiency when they return to the technology after a period of not using it. Errors, in terms of their number, severity, and ease of recovery from, are a fourth key element. Satisfaction (of users) constitutes the final element. Usability testing may also consider conformance to evidence-based standards for the design of instructional technologies, such as Mayer’s multimedia principles (Mayer 2005).

Most usability tests result in changes to the system. These should be documented (see above), and the iterative cycle of testing and improvement should continue until usability has reached an acceptable level. However, aspiring to create a flawless course may be counterproductive (Cook 2014). In many instances, a developer gains much from releasing a TEL intervention relatively early (the “minimal viable product”), carefully and deliberately evaluating the result, and then quickly making evidence-based improvements (Ries 2011).

4. Document key events during implementation and final product

Documentation continues into implementation, now focusing on objective accounts of what actually happens. This may be a normative account that checks for intentional or accidental variances from the intended execution, and the steps taken to resolve such variances, or it may be an unstructured observation of events as they occur. The final product (after any changes) should be documented carefully.

Although this Documentation activity is about objectively observed events (as compared with the subjective experiences of participants described in the next section), those doing the documenting constitute an important part of the process. Different people may have access to different information, notice different things, or interpret the same thing in different ways. A member of the instructor team would usually perform this function, but in some cases (e.g. distance learning) learners may need to be enlisted as reporters of key events.

5. Assess participant experience and satisfaction

This activity focuses on the subjective experiences of participants in a TEL activity, the value and meaning they attribute to those experiences, and their perception of the quality of the materials or tools they used. Educational evaluations commonly address this activity, which includes Kirkpatrick level 1 (Reaction) outcomes. Although experiences are also documented in other activities (as above), the focus here is on the subjective opinions of participants rather than the objective recording of events or of actual learning outcomes. Although learners are typically the focus of this evaluation activity, participants also include instructors and other
stakeholders such as administrators, support staff, and real or standardized patients; these will provide useful information for many evaluations. Most participant experiences are assessed through self-report surveys or interviews/focus groups.

6. Assess learning outcomes

As previously noted, Kirkpatrick levels 2, 3 and 4 focus on the objective outcomes of training activities (Kirkpatrick 1996). These arguably constitute the most critical element in determining the efficacy and effectiveness of an educational intervention. The link between the training event and the observed outcome becomes more difficult to establish as one progresses from level 2 to 3 to 4 (Cook & West 2013), yet the perceived meaningfulness and value also rises, emphasizing the need for thoughtful balance in selecting outcomes and the associated measurement instruments. Since the development and validation of level 2–4 outcomes have been extensively explored in other sources (Kirkpatrick 1996; Case & Swanson 2001; Cook & Beckman 2006; Kopan et al. 2009; Schuurhuis & van der Vleuten 2011; Pangaro & ten Cate 2013), we will not discuss these further.

7. Estimate cost, reusability, and sustainability

One of the most neglected (Zendejas et al. 2013), yet arguably one of the most important (Sandars 2010), aspects of evaluation focuses on the financial, personnel, facilities, and other resource costs required to develop, implement, and maintain a TEL course (Clune 2002; Hummel-Rossi & Ashdown 2002). Approaches such as Levie’s “ingredients” model (Levin 2001) and the “total cost of ownership” (TCO) model (Ellaway 2010b) estimate the true cost of an educational activity by identifying and valuing each component, including costs related to equipment and materials, licensing, personnel, facilities and infrastructure, learner expenses (such as transportation or meals), lost opportunities (e.g. lost clinical revenue), and stopping an existing activity or moving it from one medium to another (for instance when the hosting technology changes). A careful evaluation of costs can inform secondary analyses of cost-effectiveness and return on investment (ROI) (Cook 2014), and also help determine course sustainability.

Repeating and repurposing all or part of a TEL course (offering a course a second time, or reusing part of one course in another course) could potentially reduce the per-learner cost, yet in practice such economies may be less easily achieved than anticipated (Cook & Triola 2014). In addition to exploring the results of actual repeated offerings, an evaluation might seek information that predicts the potential for reuse or that indicates preparation for anticipated maintenance, updates, and integration with other educational endeavors (Glasgow et al. 1999).

Practical applications

New instruments to assess participant perceptions

To facilitate practical application of this model, we present three generic instruments for capturing the perceptions and experiences of learners and instructors following their participation in a TEL course: the Evaluation of Technology-Enhanced Learning Materials for Learner Perceptions (ETELM-LP) and for Instructor Perceptions (ETELM-IP), and a shortened version of the ETELMLP for use with very brief courses (ETELM-LP-S); see online supplemental materials. Such assessments comprise an important element of virtually all course evaluations, and as such represent a recurrent need for nearly all TEL educators. Moreover, consistent use of the same instrument across different courses, programs, institutions, and time periods will not only avert the reinvention of a new instrument for each occasion, but will also permit cross-course comparisons.

We developed the ETELm instruments both deductively and inductively. We initially identified the desired information (salient domains) and then specific items. We first identified two frameworks widely accepted in mainstream education (i.e. Quality Matters [Quality Matters Program 2014] and the SLOAN Consortium [Moore 2005]) that specify key elements of quality in an online course, and used these to deductively identify key domains and tentative items within each domain. We then inductively verified the completeness of these frameworks by identifying existing instruments and frameworks for evaluating TEL, and examining these for new domains or items not reflected in the initial two frameworks. We searched PubMed (using search strings such as “evaluat* AND (online OR Internet OR multimedia) AND (instrument OR valid*) AND medical education”) for relevant work in health professions education. We found one instrument that had been subjected to formal validation efforts (Alyusuf et al. 2013), and other articles describing general frameworks or listing specific areas for evaluation (Atkins & O’halloran 1995; Glenn 1996; Knight et al. 2004). Looking to education beyond medicine, we identified (in a non-systematic Internet search) 10 other instruments developed by academic institutions for local use (see online supplemental materials for listing). For each instrument thus identified we matched the key domains against those in the initial frameworks, and then used a thematic synthesis to identify common domains and reduce redundancy, resulting in a parsimonious list of domains that captured all of the key domains of the original instruments and frameworks. We then referred back to the original frameworks and instruments to identify specific items/questions within each domain. Finally, three non-author experts (two with expertise in online learning and one professional evaluator) reviewed the questionnaires and offered suggestions. We pilot tested the instrument with 16 students in graduate-level clinical research courses. All agreed the questions were easy to understand, applicable, and non-redundant, and 15/16 agreed that the length was appropriate.

The ETELm instruments exemplify both the principles and the approach set out in this article. They may be adapted to
context- and situation-specific needs by adding new items while retaining the generic core. We note that only some of the questions directly refer to technology issues. This reflects that while technology is an essential mediating factor in TEL, evaluation should focus on the whole educational activity.

A simple TEL evaluation plan

We acknowledge that most educators are not trained as evaluators, and may have difficulty planning a comprehensive evaluation (Oliver 2000). One potential solution to this challenge is the use of evaluation “recipes” (Harvey 1998) – brief outlines of evaluation plans designed to address a given audience and need. To this end, we propose a basic “recipe” for TEL evaluation (Box 2) that draws upon four of the seven evaluation activities defined previously, including the ETELM-LP and ETELM-IP. Although we believe this simple plan will likely meet the minimal needs of many TEL educators, we also realize that it may be inadequate for many contexts and trust that more experienced evaluators will define their own plans. The Evaluation Cookbook (available at http://www.icbl.hw.ac.uk/ltdi/cookbook/) lists additional evaluation recipes.

Moreover, consistent use of the same instrument across different courses, programs, institutions, and time periods has the potential to not only avert the reinvention of a new instrument for each occasion, but to allow cross-course comparisons.

We grounded our framework upon numerous prior models and instruments (Merisotis & Phipps 2000; Stufflebeam 2003; Knight et al. 2004; 2014; Moore 2005; Cook 2010; Ellaway 2010a; Alyusuf et al. 2015) and deliberately balanced the abstractions of academia, which foster rigor and comprehensiveness, with the practical needs of educators. Our primary intent is to guide the evaluation of specific TEL events within schools and agencies, although we also hope to favorably influence the quality of published scholarship in TEL.

In considering the applicability of this work, we note that our general model (Figure 1) and many of the specific information elements (Table 1) and evaluation activities (Table 2) are not specific to the TEL context. Moreover, the ubiquity of TEL in contemporary medical education suggests that in many instances TEL evaluation is becoming synonymous with educational evaluation in general. As such, we believe this work has application beyond TEL.

We note as a limitation that we did not develop specific tools for conducting all of the proposed evaluation activities. We further acknowledge that the instruments we developed have not yet been substantially tested in practice. Although the rigorous development process provides strong evidence of useful content, it will be important to evaluate other sources of validity evidence for ETELM scores (such as relations with other variables) (Cook & Beckman 2006). Finally, this article focuses on planning the evaluation and collecting data. Analyzing data and preparing the evaluation report are also essential steps (Oliver 2000), but beyond our present scope.
We have deliberately kept our focus fairly broad in order to meet the needs of multiple audiences and uses. It would likely be infeasible for any single evaluation to collect all of the suggested information, nor would any single audience likely to find all this information equally useful. As such, educators should use our comprehensive framework to identify the evaluation activities and specific information elements of greatest interest and utility to their target audience and context.

Glossary

Technology-Enhanced Learning (TEL): instructional events in which “technology plays a significant role in making learning more effective, efficient or enjoyable” (Goodyear & Retalis 2010, p. 8).

CIPP framework: A model for program evaluation organized around the themes of context, inputs, processes, and products (Stufflebeam 2003).

Quality Matters Program: an international organization focused on the quality assurance of online learning, see www.qualitymatters.org.

Notes on contributors

DAVID A. COOK, MD, MHPE, is Associate Director of Mayo Clinic Online Learning and Professor of Medicine and Medical Education, Mayo Clinic College of Medicine; and a Consultant in the Division of General Internal Medicine, Mayo Clinic, Rochester, Minnesota, USA.

RACHEL H. ELLAWAY, PhD, is Assistant Dean for Curriculum and Planning and Associate Professor in Human Sciences, Northern Ontario School of Medicine, Sudbury, Ontario, Canada.

Acknowledgements

We thank Marc Triola, Sören Huwendiek, and Katrín Frímannsdóttir for their comments on the participant perception instruments.

Declaration of interest: The authors report no declarations of interest.

References

Alyusuf RH, Prasad K, Abdel Satir AM, Abulkhair AA, Arora RK. 2013. Development and validation of a tool to evaluate the quality of medical education websites in pathology. J Pathol Inform 4:20.

Atkins MJ, O’Halloran C. 1995. AMEE Medical Education Guide No. 6. Evaluating multimedia applications for medical education. Med Teach 17:169–170.

Case SM, Swanson DB. 2001. Constructing written test questions for the basic and clinical sciences. 3rd ed. Philadelphia, PA: National Board of Medical Examiners.

Clune WH. 2002. Methodological strength and policy usefulness of cost-effectiveness research. In: Levin HM, Mckean PJ, editors. Cost-effectiveness and educational policy. Larchmont, NY: Eye On Education. pp 55–68.

Cook DA. 2010. Twelve tips for evaluating educational programs. Med Teach 32:296–301.

Cook DA. 2014. The value of online learning and MRI. Finding a niche for expensive technologies. Med Teach 36(11):965–972.

Cook DA, Beckman TJ. 2006. Current concepts in validity and reliability for psychometric instruments: Theory and application. Am J Med 119:166.e7–16.

Cook DA, Bondage G, Schmidt HG. 2008. Description, justification, and clarification: A framework for classifying the purposes of research in medical education. Med Educ 42:128–133.

Cook DA, Dupras DM. 2004. A practical guide to developing effective web-based learning. J Gen Intern Med 19:698–707.

Cook DA, Triola MM. 2014. What is the role of e-learning? Looking past the hype. Med Educ 48:930–937.

Cook DA, West CP. 2013. Reconsidering the focus on “outcomes research” in medical education: A cautionary note. Acad Med 88:162–167.

Ellway R. 2010a. Apples and architraves: A descriptive framework for e-learning research. Med Teach 32:95–97.

Ellway R. 2010b. Developing learning resources. In: Dornan T, Mann KV, Scherpier AJA, Spencer JA, editors. Medical education: Theory and practice. Edinburgh, UK: Churchill Livingstone. pp 265–282.

Ellway RH. 2006. Evaluating a VLE in medical education. (Doctoral dissertation). Available from Edinburgh Research Archive: http://hdl.handle.net/1842/8885.

Ellway RH. 2014. Virtual patients as activities: Exploring the research implications of an activity theoretical stance. Perspect Med Educ 3:266–277.

Ellway RH, Pusic M, Yavner S, Kulef AL. 2014a. Context matters: Emergent variability in an effectiveness trial of online teaching modules. Med Educ 48:396–396.

Ellway RH, Pusic M, Galbraith RM, Cameron T. 2014b. Developing the role of big data and analytics in health professional education. Med Teach 36:216–222.

Engeström Y. 1987. Learning by expanding: An activity-theoretical approach to developmental research. Helsinki, Finland: Orienta-Konsultit Oy.

Fink LD. 2003. Creating significant learning experiences: An integrated approach to designing college courses. San Francisco: Jossey-Bass.

Frye AW, Hemmer PA. 2012. Program evaluation models and related theories: AMEE guide no. 07. Med Teach 34:e288–299.

Glasgow RE, Vogt TM, Boles SM. 1999. Evaluating the public health impact of health promotion interventions: The RE-AIM framework. Am J Public Health 89:1322–1327.

Glenn J. 1996. A consumer-oriented model for evaluating computer-assisted instructional materials for medical education. Acad Med 71:251–255.

Goldie J. 2006. AMEE Education Guide no. 29: Evaluating educational programmes. Med Teach 28:210–224.

Goodyear P, Retalis S. 2010. Learning, technology and design. In: Goodyear P, Retalis S, editors. Technology-enhanced learning. Design patterns and pattern languages. Rotterdam, The Netherlands: Sense Publishers. pp 1–28.

Harvey J. 1998. Evaluation cookbook. Edinburgh: Learning Technology Dissemination Initiative, Heriot-Watt University.

Hummel-Rossi B, Ashdown J. 2002. The state of cost-benefit and cost-effectiveness research. In: Levin HM, Mcewan PJ, editors. Cost-effectiveness analyses in education. Rev Educ Res 72:1–30.

Kern DE, Thomas PA, Howard DM, Bass EB. 1998. Curriculum development for medical education: A six-step approach. Baltimore, MD: The Johns Hopkins University Press.

Kirkpatrick D. 1996. Revisiting Kirkpatrick’s four-level model. Train Dev 50(1):54–59.

Knight CL, Sakowksi HA, Houghton BL, Laya MB, DeWitt DE. 2004. Developing a peer review process for web-based curricula. Minting a new coin of the realm. J Gen Intern Med 19:594–598.

Korgan JR, Holmboe ES, Hauer KE. 2009. Tools for direct observation and assessment of clinical skills of medical trainees: A systematic review. JAMA 302:1316–1326.

Krug S. 2000. Don’t make me think! A common sense approach to Web usability. Indianapolis, IN: Macmillan USA.
Laurillard D. 2001. Rethinking university teaching: A conversational framework for the effective use of learning technologies. New York: Routledge Falmer.

Levin HM. 2001. Waiting for Godot: Cost-effectiveness analysis in education. New Direct Eval 90:55–68.

Mayer RE. 2005. Cognitive theory of multimedia learning. In: Mayer RE, editor. The Cambridge handbook of multimedia learning. New York: Cambridge University Press. pp 31–48.

Merisotis JP, Phipps RA. 2000. Quality on the line: Benchmarks for success in internet-based distance education. Washington, DC: Institute for Higher Education Policy.

Moore JC. 2005. The Sloan Consortium quality framework and the five pillars. Newburyport, MA: The Sloan Consortium.

Morrison GR, Ross SM, Kemp JE, Kalman H. 2010. Designing effective instruction. Hoboken, NJ: Wiley.

Nielsen J. 2012. Usability 101: Introduction to usability. Published by Nielsen Norman Group. [Accessed 24 July 2014] Available from http://www.nngroup.com/articles/usability-101-introduction-to-usability/.

Oliver M. 2000. An introduction to the evaluation of learning technology. Educ Technol Soc 3(4):20–30.

Olson CA, Shershneva MB. 2004. Setting quality standards for web-based continuing medical education. J Contin Educ Health Prof 24:100–111.

Pungaro I, ten Cate O. 2013. Frameworks for learner assessment in medicine. AMEE Guide No. 78. Med Teach 35:e1197–e1210.

Quality Matters Program. 2014. Quality matters: A national benchmark for online course design. [Accessed 24 July 2014] Available from https://www.qualitymatters.org/.

Ries E. 2011. The lean startup: How today’s entrepreneurs use continuous innovation to create radically successful businesses. New York, NY: Crown Business.

Sandars J. 2010. Cost-effective e-learning in medical education. In: Walsh K, editor. Cost-effectiveness in medical education. Oxon: Radcliffe. pp 40–47.

Schuwirth LW, van der Vleuten CP. 2011. General overview of the theories used in assessment. AMEE Guide No. 57. Med Teach 33:783–797.

Stufflebeam D. 2003. The CIPP model for evaluation. In: Kellaghan T, Stufflebeam D, editors. International handbook of educational evaluation. Vol. 9. Netherlands: Springer. pp 31–62.

Training and Executive Development Group. 2014. Training and Development: Training needs analysis. Published by U. S. Office of Personnel Management. [Accessed 24 July 2014] Available from http://www.opm.gov/policy-data-oversight/training-and-development/planning-evaluating/.

van Merriënboer JJG, Clark RE, de Croock MBM. 2002. Blueprints for complex learning: The 4C/ID Model. Educ Technol Res Dev 50(2):39–64.

Wall D. 2010. Evaluation: Improving practice, influencing policy. In: Swanwick T, editor. Understanding medical education: Evidence, theory and practice. Oxford, UK: Wiley-Blackwell. pp 356–351.

Yardley S, Dornan T. 2012. Kirkpatrick’s levels and education ‘evidence’. Med Educ 46:97–106.

Zendejas B, Wang AT, Brydges R, Hamstra SJ, Cook DA. 2013. Cost: The missing outcome in simulation-based medical education research: A systematic review. Surgery 153:160–176.

Supplementary material available online