Evaluating pain education programs: An integrated approach

Adam Dubrowski PhD^{1,2,3,4}, Marie-Paule Morin MD^{1,5,6}

A Dubrowski, M-P Morin. Evaluating pain education programs: An integrated approach. *Pain Res Manage* 2011;16(6):407-410.

Evaluation of educational programs and assessment of learning are essential to maintain high-standard health science education, which includes pain education. Current models of program evaluations applied to the education of the health professions, such as the Kirkpatrick model, are mainly outcome based. More recently, efforts have been made to examine other process-based models such as the Context Input Process Product model. The present article proposes an approach that integrates both outcome- and process-based models with models of clinical performance assessment to provide a deeper understanding of a program function. Because assessment instruments are a critical part of program evaluation, it is suggested that standardization and rigour should be used in their selection, development and adaptation. The present article suggests an alternative to currently used models in pain education evaluation.

Key Words: *Assessment; Pain education; Program evaluation*

Assessment is a process by which information is obtained relative to a known objective or goal. Assessment in health professions education (HPE) relates to a systematic process of obtaining information about an individual's change in skill level, knowledge or attitude (1). For the purpose of the present article, evaluation is regarded as a separate concept from assessment. Inherent in the idea of evaluation is 'value'. When we evaluate, we provide information that will help us or others make a judgment about a given situation. In HPE, evaluation typically pertains to the information one gains about a specific educational curriculum, program or activity. This process is important to justify the inclusion of content, such as pain, in already full curricula.

Although conceptually separate, assessments and evaluations coexist. Specifically, assessments and instruments used to conduct assessments (referred to as assessment instruments) must measure certain constructs related to the educational objectives of these programs. Thus, the choice of the assessment instruments depends on the program content. Similarly, program evaluations are dependent, to a degree, on information about outcomes. Therefore, the evaluation models are only as good as the fundamental information that the assessment instruments provide.

The present article examines issues related to educational program evaluation and learning assessment. Specifically, we attempt to integrate process- (2-4) and outcome-based (5) program evaluation models with models of performance assessment (6). This integration is driven by the need for generating information about the programs that is necessary for a better understanding of how they currently function and how they can be improved in the future. Such integration also highlights the interaction between evaluations and assessments, and suggests the importance of standardization and rigour in selection, adaptation and development of assessment instruments and evaluation methods. While the present article is not directly related to pain, an understanding of evaluation methods is important for all educators who develop and implement pain curricula.

L'évaluation des programmes d'enseignement sur la douleur : une approche intégrée

Il est essentiel d'évaluer les programmes d'enseignement et l'apprentissage pour maintenir des normes élevées d'enseignement en sciences de la santé, qui incluent l'enseignement sur la douleur. Les modèles actuels d'évaluation des programmes appliqués à l'enseignement des professionnels de la santé, tel que le modèle de Kirkpatrick, se fondent surtout sur les issues. Plus récemment, on s'est efforcé d'examiner d'autres modèles fondés sur des processus, tels que le modèle de contexte, d'apport, de processus et de produit. Le présent article expose une approche qui intègre les modèles fondés sur les issues et les processus à des modèles d'évaluation du rendement clinique afin de mieux comprendre une fonction de programme. Puisque les instruments d'évaluation sont un élément essentiel de l'évaluation des programmes, il est proposé de privilégier la normalisation et la rigueur dans leur sélection, leur élaboration et leur adaptation. Le présent article propose une autre solution aux modèles actuels d'évaluation de l'enseignement sur la douleur.

PROGRAM EVALUATION

The predominant model of program evaluation in HPE is based on outcomes, with the four-level Kirkpatrick's evaluation model being the most commonly used (5). In the present article, we encourage a different view. Specifically, we suggest that outcome-based program evaluation models may provide limited information about the evaluated program, especially if the purpose of the evaluation is to generate information that will be useful and inform decision making (7). For this purpose, we propose the use of process-oriented evaluation models, such as those developed by Stufflebeam et al (8) and Patton (3), alone or in conjunction with outcome-based models such as the one proposed by Kirkpatrick (5).

Process based: The context, input, process product model

In 1983, Stufflebeam et al (8) proposed a model designed to help evaluators generate relevant information that is useful to the decision makers (8). This model guides and assists program evaluators throughout the design, development and use of the evaluation. The approach's main focus is on the evaluation process itself; outcomes are, therefore, viewed as being part of this process. Stufflebeam et al's model is frequently referred to as the context, input, process product (CIPP) model. This acronym stands for the four different types of evaluation that represent the model's core parts – the context, the input, the process and the product (Figure 1A). This evaluation model is not to be viewed as having mandatory and fixed steps in a predefined order. Instead, Stufflebeam et al encourage educators involved in evaluation to first examine the relevance of achieving each of the four types of evaluations in a given context and then to determine the most appropriate order. Solidly grounded in principles of professional standards of evaluation (7), CIPP intends not only to provide sound evaluation of the merit and worth of a program but goes beyond, and aims at gaining a better understanding of how the program functions. Stufflebeam et al highlight the importance of process-oriented models of program evaluation with the following sentence: "Most important point of an evaluation is not to prove but to improve" (8).

¹Learning Institute; ²Research Institute, The Hospital for Sick Children; ³Department of Paediatrics, Faculty of Medicine; ⁴The Wilson Centre, Faculty of Medicine, University of Toronto; ⁵Division of Rheumatology, The Hospital for Sick Children, Toronto, Ontario; ⁶Department of Pediatrics, Division of Immunology and Rheumatology, CHU Sainte-Justine, Montreal, Quebec
Correspondence: Dr Adam Dubrowski, SickKids Learning Institute, 525 University Avenue, Room 6021, Unit 600, Toronto, Ontario M5G 2L3.
Telephone 416-813-7654, fax 416-813-6924, e-mail adam.dubrowski@utoronto.ca

Figure 1 Integrated program evaluation model. Panels **A** to **C** are the pictorial representations of (**A**) Stufflebeam et al's (8) CIPP (context, inputs, processes and products) model, (**B**) Kirkpatrick's Learning Evaluation Model (5) and (**C**) Miller's Clinical Assessment Framework (6). The CIPP is a process-based model in which outcomes or products are only part of the program evaluation. Kirkpatrick's model outcomes can help evaluators in reaching decisions about what outcomes to measure and where to measure them. Finally, according to Moore et al (14), Miller's framework can be helpful in deciding on the choice of assessments to address the specific outcomes

Context: The context of evaluation should answer the question "What needs to be done?" (9). Program evaluators, therefore, need to gather empirical data to characterize the educational environment, identify the weaknesses and the shortcomings of the current program, and uncover the problems that need to be addressed. Two recent surveys of hours allotted for pain assessment and management in prelicensure curricula indicate that content is minimal (10). The strengths of the system, such as faculty expertise and the opportunities offered, are also analyzed. The evaluators need to identify the target beneficiaries and assess their needs (8). In health profession education, beneficiaries include students, faculty members and facilitators, if involved. To conceptualize the relevant information of the educational context, evaluators can use different methods, such as interviews with program leaders or focus groups, or use of techniques, such as the Delphi method developed by the RAND Corporation in the late 1960s, as a forecasting methodology. By collecting such data, the evaluators develop a clear understanding of the current system. Ultimately, the context evaluation leads to a decision regarding the need to change the current program (8).

Input: Once the need to improve education, such as about pain, has been established, input evaluation involves the assessment of various approaches to meeting these needs and objectives. The goal is to identify and rate competing strategies. Evaluators may search and critically examine relevant literature, discuss with experts or study programs that have succeeded (8). The information retrieved should, therefore, assist in answering the question "How should it be done?" (9). Decision makers would use this information to judge whether they have identified a viable strategy, and evaluators should clarify whether the instructional system has the necessary resources to conduct the chosen approach.

Process: A key word in process evaluation is implementation. The evaluators should evaluate to what extent the program has been implemented according to the original plan, and try to identify the problems encountered. Evaluators examine and give feedback on the execution in terms of cost, efficacy and respect of the schedule. Use of periodic interviews with students, program leaders or staff, questionnaires and focus groups are different ways of conducting this type of evaluation. The information obtained answers the question "Are we doing it correctly?" (9) and may then be used by the stakeholders to refine implementation, strengthen the program design and enable improved coordination of the program's activities (8).

Product: The purpose of product evaluation is to measure to what extent the program has met the needs of the targeted beneficiaries. It is in this type of evaluation that outcomes are operationally defined

and judged against the relevant standards expected (8). Stufflebeam et al suggested that evaluators should attempt to use a combination of both qualitative and quantitative techniques in assessing outcomes to obtain a clear and comprehensive picture. In evaluating pain curricula, this can include the quantitative measures of objective structured clinical examinations (11) and/or qualitative analyses of data from focus groups. The impact of the program, its effectiveness, sustainability and transferability may also be examined.

A CIPP program evaluation example applied to medical education

CIPP is a frequently used model for educational program evaluation. It is, therefore, not surprising to encounter studies published in recent years that use the CIPP model as a framework to assess the implementation of health science educational programs. For example, Steinert et al (12) used the CIPP model to evaluate a faculty development program designed to promote the teaching of professionalism to medical students and residents. The authors conducted all four elements of the CIPP model and also provided preliminary evaluations of their program. Their faculty development initiative was, therefore, evaluated from the initial steps of its planning to the implementation and evaluation of its educational benefits and impacts. Steinert et al suggest "more rigorous evaluations of these faculty development initiatives should be conducted" (12).

Outcome-based evaluation models

Outcome-based evaluation models are best suited as research tools to provide the basis for iterative change and/or generalize findings from one program to another similar program (7). The most frequently used outcome-based evaluation model in HPE is the one proposed by Kirkpatrick (5). In its most basic version, the model uses the assessment of outcomes related to four codependent dimensions of the program to evaluate the program (Figure 1B). These dimensions include: level 1 – reactions of students about what they believed and felt about the program; level 2 – students' learning, referred to as the resulting increase in knowledge or capability; level 3 – students' behaviour, defined as the extent of behaviour and capability improvement and application to clinical practice; and, finally, level 4 – results, defined as the effects on the clinical environment, practice, or system resulting from the trainees' performance. While this model has been useful in evaluating pain curricula, the last dimension related to clinical practice transfer can be challenging to evaluate (11,13).

The four main criticisms of outcome-oriented models, such as Kirkpatrick's model, are the following: they put too much emphasis on the training program itself, rather than devoting attention to the

stakeholders; they predominantly focus on learning outcomes, specifically successes and failures, thus answering the question 'What was/was not learned?', rather than 'Why was it/was it not learned?'; they focus only on learners and not on the program objectives; and, finally, they emphasize progression to the changed behaviour in the clinical setting, without proper emphasis on integration of the program's content with the setting. Collectively, these criticisms comprise the shortcomings of outcome-based models that do not consider the context, relevant inputs and processes leading to the success of the program.

In the following section, we argue that the integration of process-based models, underpinning the implementation of a program and its success, with outcome-based models, such as the one proposed by Kirkpatrick, may lead to a more comprehensive view of how a specific program functions.

Integrated program evaluation model

Integrated models of program evaluation have previously been proposed. Most notably, Moore et al (14) speculated that the Kirkpatrick's levels associated with learning and behaviour (ie, levels 2 and 3, respectively) are closely related to Miller's framework for assessing clinical competence (6), also known as Miller's assessment pyramid (Figure 1C). At the lowest level of the pyramid is 'declarative knowledge' (knows) – the degree to which participants state what the learning activity intended them to know; followed by 'procedural knowledge' (knows how) – the degree to which participants state how to do what the learning activity intended them to know how to do; 'competence' (shows how) – the degree to which participants show, in an educational setting, how to do what the learning activity intended them to be able to do; and 'action' (does) – the degree to which participants do what the learning activity intended them to be able to do in their practices. As depicted in Figure 1, in Moore's view, level 2 of Kirkpatrick's model (ie, learning) can be evaluated using assessments related to the bottom 3 levels of Miller's framework (ie, knows, knows how and shows how); while Kirkpatrick's level 3 (behaviour) can be evaluated by looking at work-based assessments (ie, does) (14,15). Although comprehensive and useful, Moore's model is still, predominantly, an outcome-based program evaluation model. In fact, the usefulness of the model is in its integration of outcome-based program evaluation, with specific markers of changes in outcomes (assessments). However, Moore's model, as a program evaluation model, may be criticized in the same way as Kirkpatrick's model.

Our view is that for a full understanding of educational processes, an integrated process and outcome-based model of program evaluation may be useful (Figure 1). This approach involves evaluation of the context, knowledge about the inputs and processes, as well as assessment of learning outcomes (or products). In our view, this integrated model takes advantage of the strengths of the predominant, outcome-based models, such as the one proposed by Kirkpatrick (5), and addresses their criticisms with the strengths of process-based models, such as those proposed by Madaus et al (4) and Patton (3).

COMPARISON OF RIGOUR AND STANDARDIZATION

Assessment instruments

The implication of Moore's model (14) is that the selection of assessment instruments is critical for supporting program evaluation models. Therefore, assessment instruments should be regarded as tools that educators, researchers and those interested in program evaluation issues have in their 'toolbox'. For example, the Objective Structured Assessment of Technical Skills (OSATS) were developed and validated for the assessment of technical skills (16,17). The original OSATS were developed for surgical procedures and typically ignore the 'patient', who would usually be anesthetized. Therefore, their strength is in assessing performance on simulators. However, for the assessment of global clinical performances on procedures that involve an awake patient, an approach termed Integrated Procedural Performance Instrument (IPPI) has been proposed (18-21). IPPI uses scenarios involving a combination of standardized patients and

inanimate simulators, and a set of assessment instruments that are used by observers as well as the standardized patients themselves. Therefore, the assessments are geared toward technical and nontechnical performances such as communication with the patients (20).

The choices and selection of these instruments will depend on their intended use, availability and existing evidence supporting their validity. To date, however, a systematic framework for the development and validation of assessment instruments in HPE has not been proposed. We suggest an adaptation of the rigorous guidelines proposed by the Medical Outcomes Trust (MOT). The MOT encourages the following: rigorous standards for outcome measures, development of a library of instruments that meet those standards, and distribution, royalty-free, of those instruments with instructions for their use. The MOT established eight criteria to evaluate instruments:

- The Conceptual and Measurement Model: The concept to be measured needs to be defined properly and should match its intended use.
- Reliability: The degree to which the instrument is free of random error, which means free from errors in measurement caused by chance factors that influence measurement.
- Validity: The degree to which the instrument measures what it purports to measure.
- Responsiveness or sensitivity to change: An instrument's ability to detect change over time.
- Interpretability: The degree to which one can assign easily understood meaning to an instrument's score.
- Burden: Refers to the time, effort, and other demands placed on those to whom the instrument is administered (respondent burden) or on those who administer the instrument (administrative burden).
- Alternative forms of administration: Alternative means of administration include self-report, interviewer-administered, computer-assisted, etc. It is often important to know whether these modes of administration are comparable.
- Cultural and language adaptations.

In summary, assessment instruments are fundamental in providing formative and summative feedback to trainees, allowing for relative ranking of the trainees, as well as supporting evaluation models. Therefore, rigour in the choice of their adaptation or purposeful development must be exercised.

CONCLUSIONS

The present article intended to address issues related to program evaluation and assessment of learning in HPE, including pain education. Specifically, we attempted to integrate existing process-based program evaluation models (Stufflebean, Patton) and outcome-based program evaluation models (Kirkpatrick and Moore) with models of assessment of clinical performance (Miller). We propose that such an integrated program evaluation model may provide evaluators with a better understanding of the multitude of factors influencing not only the success of the program but also its sustainability, fit with other programs, transferability and ongoing improvement. Collectively, such an approach may prove to be relevant when the purpose of the evaluation is to generate information that will be useful and inform decision making (7). More specifically, we argue that one common shortcoming of outcome-based program evolution models is that while they are hierarchical, they often fail to demonstrate successful outcomes at the higher levels of behavioural change, despite the presence of well-constructed programs, sophisticated learners and success at the lower levels, such as learners' satisfaction with the program as well as acquisition of procedural and declarative knowledge (22,23).

Evaluation of the transfer of pain knowledge from academic learning to the complexity of clinical practice is challenging. Therefore, although outcomes models provide the stakeholders with a common knowledge of the program's outcomes, they need to provide more insight into processes by which a program achieves or fails to achieve

the desired outcomes (24) such as in actual patient care contexts. Finally, we suggest that assessment instruments should be regarded as fundamental ingredients in program evaluation, including those involving pain education. Therefore, rigour in the choice of their adaptation or purposeful development must be exercised.

ACKNOWLEDGEMENTS: Dr Morin's contributions were made possible by funding from the Joint Fund Award, Centre Hospitalier Universitaire Sainte-Justine, Montreal, Quebec. This work was performed at the Learning Institute, The Hospital for Sick Children, Toronto, Ontario.

REFERENCES

- Schuwirth L, van der Vleuten C. Merging views on assessment. *Med Educ* 2004;38:1208-10.
 - Patton MQ. *Utilization-focused Evaluation: The New Century Text*, 3rd edn. Thousand Oaks: Sage Publications, 1997.
 - Patton MQ. Process use as a usefulness. *New Directions for Evaluations* 2007;116:99-112.
 - Madaus GF, Scriven M, Stufflebeam DL. *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer-Nijhoff, distributors for North America Kluwer Academic Publishers; 1983.
 - Kirkpatrick DL. *Evaluating Training Programs: The Four Levels*, 2nd edn. San Francisco: Berrett-Koehler Publishers, 1998.
 - Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;65(Suppl 9):S63-7.
 - Alkin MC, Christie CA. The evaluation theory tree. In: Alkin MC, ed. *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks: Sage Publications, 2004:12-65.
 - Stufflebeam DL, Madaus GF, Kellaghan T. *Evaluation Models Viewpoints on Educational and Human Services Evaluation*, 2nd edn. Boston: Kluwer Academic, 2002:279-318.
 - Stufflebeam DL. CIPP evaluation model checklist. <www.wmich.edu/evalctr/archive_checklists/cippchecklist_mar07.pdf> (Accessed on October 11, 2010).
 - Watt-Watson J, McGillion M, Hunter J, et al. A survey of prelicensure pain curricula in health science faculties in Canadian universities. *Pain Res Manag* 2009;14:439-44.
 - Watt-Watson J, Hunter J, Pennefather P, et al. An integrated undergraduate pain curriculum, based on IASP curricula for six health science faculties. *Pain* 2004;110:140-8.
 - Steinert Y, Cruess S, Cruess R, Snell L. Faculty development for teaching and evaluating professionalism: From programme design to curriculum change. *Med Educ* 2005;39:127-36.
 - Hunter J, Watt-Watson J, McGillion M, et al. An interfaculty pain curriculum: Lessons learned from six years experience. *Pain* 2008;140:74-86.
 - 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.
 - Norcini JJ. Peer assessment of competence. *Med Educ* 2003;37:539-43.
 - Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med* 1996;71:1363-5.
 - Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. *Am J Surg* 1997;173:226-30.
 - Moulton CA, Tabak D, Kneebone R, Nestel D, MacRae H, LeBlanc VR. Teaching communication skills using the integrated procedural performance instrument (IPPI): A randomized controlled trial. *Am J Surg* 2009;197:113-8.
 - LeBlanc VR, Tabak D, Kneebone R, Nestel D, MacRae H, Moulton CA. Psychometric properties of an integrated assessment of technical and communication skills. *Am J Surg* 2009;197:96-101.
 - Kneebone R, Bello F, Nestel D, Yadollahi F, Darzi A. Training and assessment of procedural skills in context using an Integrated Procedural Performance Instrument (IPPI). *Stud Health Technol Inform* 2007;125:229-31.
 - Kneebone R, Bello F, Nestel D, et al. Learner-centred feedback using remote assessment of clinical procedures. *Med Teach* 2008;30:795-801.
 - Mansouri M, Lockyer J. A meta-analysis of continuing medical education effectiveness. *J Contin Educ Health Prof* 2007;27:6-15.
 - Noe R, Schmitt N. The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology* 1986;39:497-523. <<http://onlinelibrary.wiley.com/journal/10.1111/%28ISSN%291744-6570>> (Accessed on October 7, 2011).
 - Chen H, Rossi P. The theory driven approach to validity. *Evaluation and Program Planning* 1987;10:95-103.
-