

# Advancing Faculty Development in Medical Education: A Systematic Review

Karen Leslie, MD, MEd, Lindsay Baker, MEd, Eileen Egan-Lee, MEd, Martina Esdaile, MA, and Scott Reeves, PhD, MSc

## Abstract

### Purpose

To (1) provide a detailed account of the nature and scope of faculty development (FD) programs in medical education, (2) assess the quality of FD studies, and (3) identify in what areas and through what means future research can purposefully build on existing knowledge.

### Method

The authors searched MEDLINE, CINAHL, and ERIC for articles reporting evaluations of FD initiatives published between 1989 and 2010. They applied standard systematic review procedures for sifting abstracts, scrutinizing full texts, and abstracting data, including program characteristics, evaluation methods, and outcomes. They used a

modified Kirkpatrick model to guide their data abstraction.

### Results

The authors included 22 articles reporting on 21 studies in their review. The most common program characteristics included a series/longitudinal format, intended for individuals, and offered to physicians only. Although the most common aim was to improve teaching effectiveness, several programs had multiple aims, including scholarship and leadership. Program evaluation focused on quantitative approaches. A number of studies employed longitudinal designs and included some follow-up component. Surveys were the

most popular data collection method, participants the most common data source, and self-reported behavior changes the most commonly reported outcome.

### Conclusions

Although the authors' findings showed some recent expansion in the scope of the FD literature, they also highlighted areas that require further focus and growth. Future research should employ more rigorous evaluation methods, explore the role of interprofessional teams and communities of practice in the workplace, and address how different organizational and contextual factors shape the success of FD programs.

Throughout the last few decades, both health care delivery and medical education have undergone extensive changes.<sup>1</sup> Such changes have resulted in increasing demands on faculty to be "creative and effective teachers, successful researchers, and productive clinicians."<sup>2</sup> Thus, to meet these demands, faculty have had to acquire new knowledge, skills, and abilities in a relatively short period of time.<sup>2,3</sup> Faculty development (FD) is recognized by many medical education organizations as an essential support framework provided to faculty members to assist them in responding to the challenges of their multiple roles and evolving responsibilities.

Stritter<sup>4</sup> initially conceptualized FD as strategies to improve faculty members' teaching performance.

Please see the end of this article for information about the authors.

Correspondence should be addressed to Dr. Leslie, Centre for Faculty Development, St. Michael's Hospital, 30 Bond St., Toronto ON, M5B 1W8; telephone: (416) 864-6060 ext. 77412; fax: (416) 864-5929; e-mail: lesliek@smh.ca.

*Acad Med.* 2013;88:1038–1045.  
First published online May 22, 2013  
doi: 10.1097/ACM.0b013e318294fd29

Subsequent reviews<sup>5,6</sup> have called for a broader definition of FD, based on the expanding scope of faculty roles, including administration, scholarship, and leadership. However, a series of surveys<sup>1,7,8</sup> conducted over the past 30 years, which described and tracked changes in FD programs in Canadian medical schools, indicate that although such programs have begun to address the breadth of faculty members' needs, their focus remains on strategies to improve teaching performance.

In recent years, the FD literature has grown. As a result, we have a clearer understanding of the ways in which this type of education has been delivered and the kinds of outcomes that have been achieved. We know, for example, that FD initiatives can be effective in improving knowledge and self-perceived changes in teaching behavior.<sup>9,10</sup> However, despite these advances in our understanding of FD, there has been a limited amount of work that synthesizes this growing volume of studies. To date, this work has focused on the assessment of the quality of a particular aspect of a FD activity, such as teaching improvement or mentorship,<sup>11,12</sup> and does not provide

readers with a complete understanding of the full range of FD activities and interventions.

The overall aim of this systematic review was to describe the range of FD activities that have been developed and implemented within medical education and to assess the current quality of the evaluation of these initiatives. Our specific objectives were threefold: (1) to provide an account of the nature and scope of FD programs, (2) to provide an assessment of the quality of FD studies, and (3) to identify in what areas and through what means future research can purposefully build on existing knowledge. Understanding the nature of FD studies as well as their outcomes has a number of far-reaching implications for medical schools in terms of how they design, implement, and evaluate FD programs.

### Method

#### Eligibility criteria

We included articles that reported program evaluations of FD initiatives for both basic science and clinical faculty in academic medicine. Although we included programs with participants

from across the health professions, all had to include physicians to be considered eligible for review. We included only peer-reviewed articles published between January 1989 and December 2010. Although we did not limit our search by language or country of practice, we did limit our review to articles published in English.

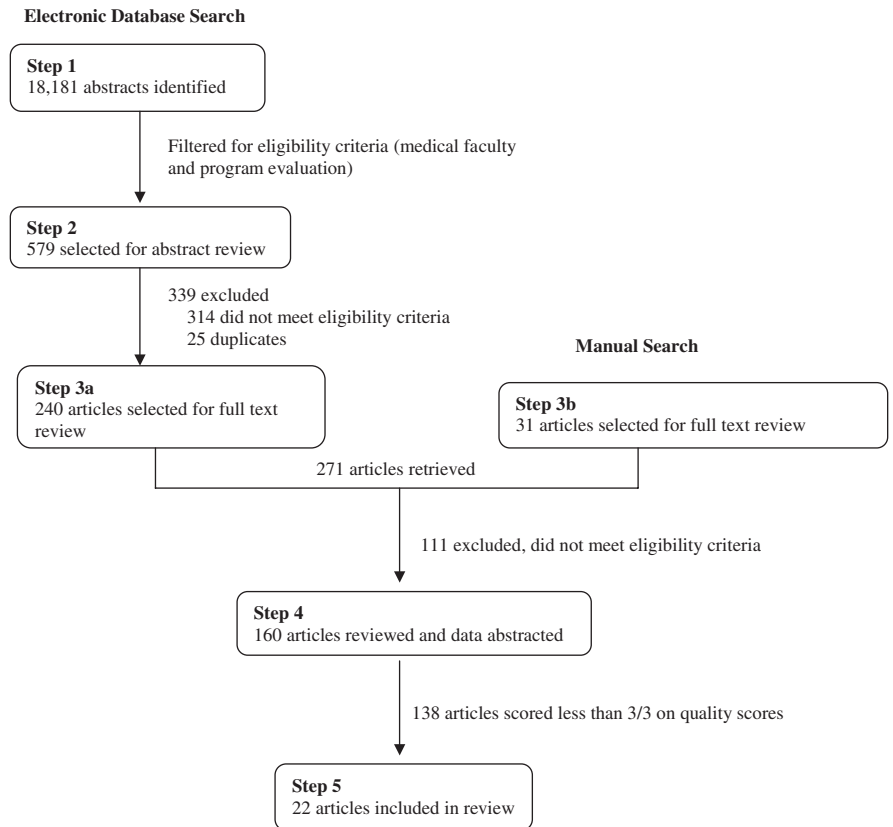
We excluded articles that had been included in previous FD reviews (e.g., by Steinert and colleagues<sup>12</sup>) to avoid repetition and to build on their findings.

### Search strategies and selection methods

We used two approaches to locate articles for inclusion in our review. First, we searched for relevant articles published in the past 21 years (January 1989 to December 2010) using the electronic databases MEDLINE, CINAHL, and ERIC. We chose these three databases because they span the health professions. We sequentially searched each of the databases to account for the substantial overlap between them. We searched the titles and abstracts of articles using search terms that were a combination of controlled vocabulary from the thesaurus subject headings of the three databases and free-text key words. We combined our search terms for FD (faculty development, staff development, professional development, professional training) with those for evaluation (evaluation, evaluation studies, program evaluation, program effectiveness, effectiveness, efficacy, impact, outcome assessment). In addition, we conducted a search for any systematic reviews and meta-analyses in the FD literature by combining the search terms for FD as listed above with the relevant publication type (systematic review or meta-analysis).

Second, we conducted manual searches of three leading medical education journals that publish articles on FD initiatives (*Academic Medicine*, *Medical Education*, and *Medical Teacher*) for the same period (January 1989 to December 2010).

We applied standard systematic review procedures for sifting abstracts, scrutinizing full papers, and abstracting data (see Figure 1).<sup>13</sup> At least two members of the team read each full text article and abstracted the pertinent data. A third member of the team was consulted if a difference of opinion arose.



**Figure 1** Flowchart of literature search and study selection process from a systematic review of the literature on faculty development programs published between January 1989 and December 2010.

Once this process was complete, a fourth member of the team independently abstracted the pertinent data from the included studies.

### Data abstraction, analysis, and synthesis

We developed a data extraction sheet through iterative testing and revision. Variables coded included characteristics of the FD initiative (e.g., program type, duration) and outcomes. We were also interested in the robustness of the methods and coded for characteristics of the evaluation (e.g., research design and data collection).

To guide our abstraction of the different outcomes of the FD programs, we used Kirkpatrick's<sup>14</sup> model of educational outcomes, which offers a useful four-point typology of educational outcomes (learner reaction, acquisition of learning, behavioral change, and changes in organizational practice). Building on Barr and colleagues<sup>15</sup> and Steinert and colleagues<sup>12</sup> use of the Kirkpatrick model, we further modified our list of outcomes to the following seven categories: (1) learner reaction (level 1), (2) modification

of attitudes/perceptions (level 2a), (3) acquisition of knowledge/skills (level 2b), (4) behavioral change (level 3), (5) changes in organizational practice (level 4a), (6) benefits to students/residents (level 4b), and (7) benefits to patients/communities (level 4c).

To facilitate comparisons and summaries of data, we further condensed the abstraction sheet into SPSS version 19.0 (Armonk, New York, IBM Corp.). We input all data into the SPSS file and calculated frequencies and cross-tabulations to produce a synthesized descriptive account of the articles.

### Assessing the quality

To identify studies that were more robust, we included an extra step in the abstraction process. We calculated scores (out of five points) along two dimensions—the quality of the study and the quality of the information provided in the article.<sup>16</sup> Only articles that attained at least three points on both dimensions were eligible for inclusion in our review. The quality of the study score reflected the quality of the design and execution of the

study—for example, a good fit between the methodological approach and research questions, attention to ethical concerns, adequate recruitment and retention of participants, and appropriate analysis. Thus, when the research aims/questions were oriented toward quantifying outcomes, a well-designed, well-conducted pre/post study could potentially score a 5 for methodological quality. Similarly, when the research aims/questions were oriented toward understanding processes, a high-quality ethnography study also could score a 5. Studies that were competently conducted with clear objectives and inclusion/exclusion criteria but lacked sufficient detail concerning data analysis or attention to issues of bias received midrange scores. Studies with weak designs in relation to research questions scored a 1—for example, postintervention studies or descriptive studies that lacked detail about research aims/questions and data collection and analysis and failed to consider issues of bias.

The quality of the information score took into account a number of factors, such as whether a detailed description of the FD initiative was provided, whether a clear rationale for the evaluation was given, and whether the analysis was described in sufficient detail.

## Results

In total, our database and manual searches produced 18,212 references related to FD (see Figure 1). Of those, 160 met our eligibility criteria. Of those, 22 met our quality criteria and were included in our final review. We present our results in three sections—FD initiatives, evaluation approaches, and reported outcomes. Appendix 1 provides a summary of the key findings from the 22 articles included in our review.

### FD initiatives

Of the 22 articles, 15 reported on FD initiatives that took place in the United States,<sup>17–31</sup> 3 in Canada,<sup>32–34</sup> and 1 in each of Israel,<sup>35</sup> Sweden,<sup>36</sup> and Germany.<sup>37</sup> One described an international collaboration between the United States, Canada, and Puerto Rico.<sup>38</sup> Nearly all the articles (n = 21) were published from 2001 to 2010 (one was published in 1990<sup>23</sup>). Although the majority of articles used the term *faculty development* (n = 20), other

terms used included *staff development*,<sup>36</sup> *teaching workshop*,<sup>32</sup> *consulting program*,<sup>38</sup> and *tutor-training program*.<sup>27</sup> However, only two articles provided a definition for the term they used.<sup>23,33</sup>

Two articles described the same program<sup>30,31</sup>; therefore, we reviewed 22 articles but 21 programs. Of the 21 programs, the most common format described was series/longitudinal (n = 12).<sup>18–21,24,25,28–31,36–38</sup> These programs were either a series of workshops or a longitudinal program that participants attended over a prolonged length of time (ranging from 10 days to 2 years). Four programs were single workshops (one day or less),<sup>17,27,33,35</sup> 2 were short courses (less than one week),<sup>22,32</sup> and 1 was a fellowship program (1 year).<sup>23</sup> Two programs did not fit into these categories—the first involved observations of workplace teaching followed by feedback,<sup>26</sup> and the second involved a combination of a workshop, a series of peer writing groups, and independent study.<sup>34</sup>

The majority of programs were intended for individual learners (n = 19)<sup>17–21,23–37</sup> rather than teams (n = 2).<sup>22,38</sup> Fifteen were intended for physicians only,<sup>17–20,22–24,26–29,33,35,37,38</sup> whereas 6 included a mix of health professionals (including nursing, pharmacy, public health, dentistry, basic science, and rehabilitation science).<sup>21,25,30–32,34,36</sup> The scope of the programs ranged from local (n = 11) to national (n = 9) to international (n = 1). The articles did not explicitly discuss a theoretical framework for the FD activities, with the exception of Sullivan and colleagues,<sup>30,31</sup> which mentioned the use of adult learning theories in the instructional design of their program.

Many of the included studies had multiple aims. The most common program aim was to improve teaching effectiveness; 15 of the 21 programs included this goal as one of their primary objectives.<sup>17–20,23–27,30–33,35–37</sup> The second most common program aim (n = 8) was scholarship,<sup>21,23–25,28,33,34,37</sup> which encompassed such activities as curriculum design and the development of research skills. Four programs had the development of faculty developers as an objective<sup>22,28,29,38</sup>; that is, participants attended the initiative to become faculty developers themselves and to implement FD initiatives at their home institutions.

In addition, 4 programs described career development as an objective<sup>20,23–25</sup>; they aimed to nurture participants' professional effectiveness, professional academic skills, career management, and administration skills. Finally, 3 programs noted leadership as an aim,<sup>24,28,29</sup> including enhancing participants' ability to understand and influence change in their local setting, gaining leadership skills, and creating leadership focused on changing culture.

### Evaluation approaches

Table 1 provides a summary of the evaluation designs, data collection methods, and data analysis approaches employed by the studies in our review. Although 8 of the studies reported the use of mixed methods<sup>19,25,26,28,30,34,37,38</sup> and 2 studies employed qualitative methods only,<sup>29,36</sup> the focus remained predominantly on quantitative approaches, with 12 studies employing only quantitative methods.<sup>17,18,20–24,27,31–33,35</sup> Only 4 studies mentioned a theoretical or conceptual framework for the evaluation design.<sup>24,25,33,37</sup>

A number of studies employed longitudinal designs—6 with more than three data collection points over time.<sup>22,24,27,29,32,34</sup> Fifteen studies included some follow-up component, ranging from 2 months to 13 years post intervention. In addition, 9 studies included a control or comparison group in their design.<sup>17,19–21,23,26,27,32,35</sup>

Although 9 studies used more than one method of data collection, 13 studies relied on only one data collection method. Not surprisingly, surveys were the most popular method to collect data (n = 18).<sup>17,19–24,27–35,37,38</sup> These ranged from complex research instruments to “happy sheets,” which gathered participants' immediate reactions to the program. Six of these studies used a previously validated instrument. In addition, 3 evaluations analyzed data from interviews<sup>25,36,37</sup> and focus groups,<sup>37</sup> which were recorded and transcribed, 3 collected observational data,<sup>26,34,38</sup> and 3 analyzed the curriculum vitae of participants.<sup>24,25,34</sup> Other methods described included analyzing teaching scores, student marks, and progress reports.

Half of the studies used more than one data source (n = 11).<sup>17,19–21,23,26,29,32,34,35,38</sup>

**Table 1**  
**Evaluation Approaches Identified in a Systematic Review of the Literature on Faculty Development Programs Published Between January 1989 and December 2010**

Approach	No. (% of 22)
<b>Evaluation design*</b>	
Longitudinal	6 (27)
Post and follow-up	3 (14)
Pre, during, and post	2 (9)
Pre and follow-up	2 (9)
During and post	2 (9)
Pre and post	2 (9)
Retro-pre/post and follow-up	1 (4.5)
Pre, retro-pre/post, and follow-up	1 (4.5)
Pre, post, and follow-up	1 (4.5)
Post only	1 (4.5)
Follow-up only	1 (4.5)
<b>Data collection method</b>	
Survey, unvalidated	12 (55)
Other (i.e., teaching scores, students' marks)	7 (32)
Survey, validated	6 (27)
Observations	3 (14)
Interviews, recorded and transcribed	3 (14)
Curriculum vitae	3 (14)
Focus groups	1 (4.5)
<b>Data sources†</b>	
Participants	21 (95)
Comparison group	8 (36)
Students	6 (27)
Facilitators	2 (9)
Peers/colleagues	0 (0)

\*Indicates when data were collected: longitudinal (collected at more than three points); pre (collected immediately before initiative); post (regarding postinitiative attitudes and knowledge, collected immediately following initiative); retro-pre (regarding preinitiative attitudes and knowledge, collected immediately following initiative); and follow-up (collected at a point beyond the immediate completion of the initiative).

†Indicates the individuals who contributed to the study.

In general, participants were the most common source of data ( $n = 21$ ). However, at times, data collected from participants were augmented by data gathered from comparison groups ( $n = 8$ ), students ( $n = 6$ ), and facilitators ( $n = 2$ ).

### Reported outcomes

**Level 1: reaction.** Nine studies assessed outcomes at this level, which included participants' satisfaction, perception of program usefulness and acceptability, and value of the activity.<sup>20,24,28,30,32-34,37,38</sup> Participants' reactions were usually measured with a survey immediately following the program.

### Level 2a: attitudes/perceptions.

Fourteen studies addressed participants' attitudes, which included motivation,

self-confidence, enthusiasm, and conceptions of teaching and learning.<sup>17,19-21,23,25,26,28,30-33,36,37</sup> This outcome was largely self-reported ( $n = 12$ ); however, students and residents observed and reported shifts in faculty member participants' attitudes in 2 studies.<sup>26,32</sup> This outcome also was most often measured using surveys ( $n = 9$ ).<sup>17,20,21,23,28,30-33</sup> In addition, 6 of these studies recruited a comparison group of faculty to either fill out the survey themselves or to have their students/residents complete it with them in mind. Finally, in 3 studies, interviews were used to collect data about participants' attitudes.<sup>25,36,37</sup>

**Level 2b: knowledge/skills.** Sixteen studies evaluated outcomes related to participants'

knowledge and skills.<sup>19-21,24-33,35-37</sup> Although self-reported data were most common ( $n = 12$ ), 5 studies presented data related to participants' knowledge and skills as observed by others (e.g., expert medical educators). Surveys were the most common data collection method, used in 11 of the 16 studies.<sup>20,21,23-35</sup> In addition, interviews were employed in 3 studies.<sup>25,36,37</sup>

**Level 3: behavior.** By far, the most commonly reported outcome was participants' behavior change, measured in 21 of the 22 studies.<sup>17-19,21-38</sup> Behaviors measured included delivery of workshops, educational practices and teaching skills, and research productivity. Fourteen studies presented self-reported behavior outcomes, whereas 7 reported participants' behaviors as observed by others (e.g., students). Two studies included both self-reported and non-self-reported outcomes.<sup>19,29</sup> In comparison with the other reported outcomes, a variety of methods were used to gather data about participants' behavior change, including surveys ( $n = 16$ )<sup>17,19,21-23,27-35,37,38</sup> and interviews with participants ( $n = 3$ ),<sup>25,36,37</sup> the collection of observational and video data ( $n = 2$ ),<sup>18,26</sup> the analysis of curriculum vitae to track career achievements ( $n = 3$ ),<sup>24,25,34</sup> and the analysis of narratives written by participants to illustrate the influences of the FD process on their behavior ( $n = 1$ ).<sup>19</sup>

### Level 4a: organizational practice.

These outcomes measured changes that affected the organization in some way, such as the development of new programs or new curricula; the retention of faculty; new hires; and culture changes. Organizational changes were reported in 9 studies<sup>19,22,24,26,28,29,31,33,38</sup> and were mostly captured by self-reported follow-up surveys (up to 24 months after participation in the FD initiative) and progress reports submitted by participants.

**Level 4b: student benefit.** Three studies assessed the benefits to students of FD programs.<sup>19,27,29</sup> All three reported the results of surveys completed by individuals other than the FD program participants.

**Level 4c: patient benefit.** Two studies included the self-reported benefits to patients.<sup>31,37</sup> Participants completed

surveys about how the changes they made in their clinical practices as a result of the FD program affected the quality of their patient care.

## Discussion

This review provides a detailed account of the current landscape of the FD literature. It builds on the findings of previous reviews in the field<sup>11,12</sup> and sets the stage for future considerations regarding empirical work in FD. In their review focusing on improvements in teaching effectiveness, Steinert and colleagues<sup>12</sup> called for the use of rigorous research methods employed in a systematic fashion and embedded within a theoretical or conceptual framework. They highlighted the need for the use of multiple sources of data and validated instruments, the evaluation of change over time, and the assessment of organizational/institutional impact. Similarly, in their review focusing on mentorship, Sambunjak and colleagues<sup>11</sup> highlighted the poor quality of evidence in the literature. They recommended that future research employ more robust study designs performed across multiple sites that addressed contextual issues beyond individual performance.

Our review found some expansion in both the scope of FD initiatives in recent years and the evaluation methods employed by researchers, compared with the findings of Steinert and colleagues<sup>12</sup> and Sambunjak and colleagues.<sup>11</sup> For example, our findings suggest that FD programs are beginning to move away from a focus on teaching performance alone toward a variety of objectives, often within the same program. Programs are increasingly aiming to assist faculty with their scholarship, leadership, and career development needs, in addition to their teaching skills. This shift may mirror the evolving needs of faculty in response to the changing landscape of medical education and the health care system.<sup>2,3</sup> Interestingly, the development of faculty developers was an aim in several of the programs, addressing the need to extend knowledge about how best to build capacity.

The most common format for FD initiatives was a series or longitudinal program (see Appendix 1). Although these initiatives often were simply a series of workshops that participants

could attend, that we found them to be the most common format indicates that the designers of FD initiatives are moving away from the traditional format of single, one-time workshops. This shift may indicate an acknowledgement by leaders in the field that prolonged exposure (with the opportunity for the application of and reflection on learning and for reflection on practice) is often necessary for change in practice.<sup>39</sup> The majority of the FD programs, however, remained narrow in scope. They were largely focused on individuals as opposed to teams, were mostly offered to a single profession (physicians), at single sites, and to local participants only. Moreover, the development of FD initiatives appears to remain largely atheoretical, with few studies identifying a conceptual framework that informed its design.

With respect to program evaluation methods, all but one of the articles included in our review were published in the last decade (see Appendix 1), suggesting that the caliber of evaluation work has improved in recent years. Although this change is promising, there is still room for improvement. A small number of studies based their evaluations on a theoretical or conceptual framework. In addition, although qualitative approaches and mixed-method approaches to program evaluation are becoming more prevalent, the majority of studies only employed quantitative methods (see Appendix 1). This practice may be due to resource issues, including time and money, but it also may reflect medical educators' traditional preference for undertaking quantitative research work. We did, however, notice a shift from postintervention studies toward longitudinal evaluations. In addition, a growing number of studies are employing control or comparison groups in their designs. These practices indicate that quality evaluations are becoming more rigorous.

Our findings also illustrate that researchers continue to rely on a single method of data collection (see Appendix 1). Although some studies used interviews, observations, and curriculum vitae analysis, the most common form of data collection was the use of unvalidated surveys. Similarly, despite an increasing number of studies employing more than one data source (including students and FD program facilitators), program

participants remained the predominant source of data. This reliance on self-reported data is a common thread in the FD literature over the years.<sup>12</sup>

Finally, our findings indicate that the most common outcomes measured included participants' self-reported behavior changes, acquisition of knowledge and skills, and changes in attitudes and perceptions (see Appendix 1). Although this shift from relying on reaction outcomes is a welcome change, little focus has been placed on the educational process or the interplay of contextual factors that affect the success of FD. Perhaps, as O'Sullivan and Irby<sup>40</sup> suggest, it is time to move away from the traditional linear model of FD research that focuses on the individual participant. They offer instead a new model that is more cyclical in nature and that focuses on the interaction between the FD community and the workplace community.

Our review has several limitations. First, we used bibliographic databases to identify the potential articles for our review. Although doing so provided an efficient source of material, it also limited us to the resources included in the databases we chose to search (MEDLINE, CINAHL, ERIC). Second, the geographic distribution of the journals included in these databases likely is skewed toward those published in North America, which also could have limited the articles we included in our review. Finally, we noted earlier the growth in the number of FD publications over the last decade (2001–2010). This trend likely continued through to the present day, but our data set is bounded by the dates of our search, and thus we could not draw such conclusions about studies published after 2010.

On the basis of our findings, we propose the following recommendations for future FD research. First, researchers must continue the trend toward more rigorous approaches to program evaluation. The growing use of mixed methods should be encouraged because such approaches provide for comprehensive and robust studies that produce rich data. Combining both qualitative and quantitative perspectives allows researchers to generate findings that focus on both the teaching processes and the outcomes of those processes.

Related to this trend is the need for the use of theoretical frameworks in designing evaluation studies. Grounding such studies in the broader literature is necessary if FD scholarship is to engage in dialogue and align with the health professions education research community as a whole. Second, researchers must expand beyond studying solely participants' outcomes; they also must include multiple sources of data. For example, only a small percentage of the studies included in our review employed facilitators as data sources, and those that did usually used them to provide their perceptions of the changes in participants rather than to share their own experiences. Others support our call to explore further the role that facilitators can play in measuring the success of FD programs.<sup>40,41</sup> Similarly, FD research currently overlooks the role of interprofessional teams and communities of practice in the workplace. Furthering this line of research would put us a step closer to understanding how behavior change occurs within the practice environment. Finally, the bulk of FD evaluations are completed at a single institution, which limits the inferences one can draw from such studies. More multisite studies are needed to produce more compelling empirical research. Multisite studies also would explore the complex ways in which different organizational and contextual factors shape the success of FD programs.

In conclusion, our findings demonstrate a continued expansion in the scope of the FD literature. Although our review identified a number of improvements in the design, delivery, and evaluation of FD activities, it also highlighted areas that require further development. Future FD work should focus on the use of interprofessional education, the efficacy of work-based FD activities, and the effects of different organizational and contextual factors. Future research also should employ more rigorous evaluation methods to measure the impact of FD programs.

*Acknowledgments:* The authors wish to thank Rita Shaughnessy for her help developing the search strategy and identifying relevant abstracts.

*Funding/Support:* A Faculty Development Grant from the Royal College of Physicians and Surgeons of Canada funded this study.

*Other disclosure:* None.

*Ethical approval:* Not applicable.

*Previous presentations:* Select results were presented at the 2011 International Conference on Faculty Development (May 2011, Toronto, Ontario, Canada).

**Dr. Leslie** is director, Centre for Faculty Development, University of Toronto, Li Ka Shing International Healthcare Education Centre, St. Michael's Hospital, and associate professor, Department of Pediatrics, University of Toronto, Faculty of Medicine, Toronto, Ontario, Canada.

**Ms. Baker** is education research associate, Centre for Faculty Development, University of Toronto, Li Ka Shing International Healthcare Education Centre, St. Michael's Hospital, Toronto, Ontario, Canada.

**Ms. Egan-Lee** is education research associate, Centre for Faculty Development, University of Toronto, Li Ka Shing International Healthcare Education Centre, St. Michael's Hospital, Toronto, Ontario, Canada.

**Ms. Esdaile** is research assistant, Centre for Faculty Development, University of Toronto, Li Ka Shing International Healthcare Education Centre, St. Michael's Hospital, Toronto, Ontario, Canada.

**Dr. Reeves** is director, Center for Innovation in Interprofessional Education, and professor, Department of Social and Behavioral Sciences and Department of Medicine, University of California, San Francisco, and affiliate scientist, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Ontario, Canada.

## References

- McLeod PJ, Steinert Y, Nasmith L, Conochie L. Faculty development in Canadian medical schools: A 10-year update. *CMAJ*. 1997;156:1419–1423.
- Wilkerson L, Irby DM. Strategies for improving teaching practices: A comprehensive approach to faculty development. *Acad Med*. 1998;73:387–396.
- Ullian JA, Stritter FT. Faculty development in medical education, with implications for continuing medical education. *J Contin Educ Health Prof*. 1996;16:181–190.
- Stritter FT. Faculty evaluation and development. In: McGuire CH, Foley RP, Gorr A, Richards RW, eds. *Handbook of Health Professions Education*. San Francisco, Calif: Jossey-Bass; 1983:294–318.
- Bland CJ, Schmitz CC. Characteristics of the successful researcher and implications for faculty development. *J Med Educ*. 1986;61:22–31.
- Sheets KJ, Schwenk TL. Faculty development for family medicine educators: An agenda for future activities. *Teach Learn Med*. 1990;2:141–148.
- McLeod PJ. Faculty development practices in Canadian medical schools. *CMAJ*. 1987;136:709–712.
- McLeod PJ, Steinert Y. The evolution of faculty development in Canada since the 1980s: Coming of age or time for a change? *Med Teach*. 2010;32:e31–e35.
- Hewson MG, Copeland HL, Fishleder AJ. What's the use of faculty development? Program evaluation using retrospective self-assessments and independent performance ratings. *Teach Learn Med*. 2001;13:153–160.
- Skeff KM, Stratos GA, Bergen MR, Sampson K, Deutsch SL. Regional teaching improvement programs for community-based teachers. *Am J Med*. 1999;106:76–80.
- Sambunjak D, Straus SE, Marusic A. Mentoring in academic medicine: A systematic review. *JAMA*. 2006;296:1103–1115.
- Steinert Y, Mann K, Centeno A, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME guide no. 8. *Med Teach*. 2006;28:497–526.
- Petticrew M, Roberts H. *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford, UK: Blackwell; 2006.
- Kirkpatrick DL. Evaluation of training. In: Craig RL, Bittel LR, eds. *Training and Development Handbook*. New York, NY: McGraw-Hill; 1967.
- Barr H, Koppel I, Reeves S, Hammick M, Freeth D. *Effective Interprofessional Education: Argument, Assumption, and Evidence*. Oxford, UK: Wiley-Blackwell; 2005.
- Huwiler-Müntener K, Jüni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA*. 2002;287:2801–2804.
- Alford DP, Richardson JM, Chapman SE, Dubé CE, Schadt RW, Saitz R. A Web-based Alcohol Clinical Training (ACT) curriculum: Is in-person faculty development necessary to affect teaching? *BMC Med Educ*. 2008;8:11.
- Berbano EP, Browning R, Pangaro L, Jackson JL. The impact of the Stanford Faculty Development Program on ambulatory teaching behavior. *J Gen Intern Med*. 2006;21:430–434.
- Branch WT Jr, Frankel R, Gracey CF, et al. A good clinician and a caring person: Longitudinal faculty development and the enhancement of the human dimensions of care. *Acad Med*. 2009;84:117–125.
- Cole KA, Barker LR, Kolodner K, Williamson P, Wright SM, Kern DE. Faculty development in teaching skills: An intensive longitudinal model. *Acad Med*. 2004;79:469–480.
- Gozu A, Windish DM, Knight AM, et al. Long-term follow-up of a 10-month programme in curriculum development for medical educators: A cohort study. *Med Educ*. 2008;42:684–692.
- Houston TK, Clark JM, Levine RB, et al. Outcomes of a national faculty development program in teaching skills: Prospective follow-up of 110 medicine faculty development teams. *J Gen Intern Med*. 2004;19:1220–1227.
- McGaghie WC, Bogdewic S, Reid A, Arndt JE, Stritter FT, Frey JJ. Outcomes of a faculty development fellowship in family medicine. *Fam Med*. 1990;22:196–200.
- Morzinski JA, Simpson DE. Outcomes of a comprehensive faculty development program for local, full-time faculty. *Fam Med*. 2003;35:434–439.
- Moses AS, Skinner DH, Hicks E, O'Sullivan PS. Developing an educator network: The effect of a teaching scholars program in the health professions on networking and productivity. *Teach Learn Med*. 2009;21:175–179.
- Regan-Smith M, Hirschmann K, Jobst W. Direct observation of faculty with feedback: An effective means of improving patient-centered and learner-centered teaching skills. *Teach Learn Med*. 2007;19:278–286.

- 27 Shields HM, Guss D, Somers SC, et al. A faculty development program to train tutors to be discussion leaders rather than facilitators. *Acad Med.* 2007;82:486–492.
- 28 Simpson DE, Bragg D, Biernat K, Treat R. Outcomes results from the evaluation of the APA/HRSA Faculty Scholars Program. *Ambul Pediatr.* 2004;4(1 suppl):103–112.
- 29 Stratos GA, Katz S, Bergen MR, Hallenbeck J. Faculty development in end-of-life care: Evaluation of a national train-the-trainer program. *Acad Med.* 2006;81:1000–1007.
- 30 Sullivan AM, Lakoma MD, Billings JA, Peters AS, Block SD; PCEP Core Faculty. Teaching and learning end-of-life care: Evaluation of a faculty development program in palliative care. *Acad Med.* 2005;80:657–668.
- 31 Sullivan AM, Lakoma MD, Billings JA, Peters AS, Block SD; PCEP Core Faculty. Creating enduring change: Demonstrating the long-term impact of a faculty development program in palliative care. *J Gen Intern Med.* 2006;21:907–914.
- 32 Pandachuck K, Harley D, Cook D. Effectiveness of a brief workshop designed to improve teaching performance at the University of Alberta. *Acad Med.* 2004;79:798–804.
- 33 Steinert Y, Cruess S, Cruess R, Snell L. Faculty development for teaching and evaluating professionalism: From programme design to curriculum change. *Med Educ.* 2005;39:127–136.
- 34 Steinert Y, McLeod PJ, Liben S, Snell L. Writing for publication in medical education: The benefits of a faculty development workshop and peer writing group. *Med Teach.* 2008;30:e280–e285.
- 35 Notzer N, Abramovitz R. Can brief workshops improve clinical instruction? *Med Educ.* 2008;42:152–156.
- 36 Weurlander M, Stenfors-Hayes T. Developing medical teachers' thinking and practice: Impact of a staff development course. *Higher Educ Res Dev.* 2008;27:143–153.
- 37 Herrmann M, Lichte T, Von Unger H, et al. Faculty development in general practice in Germany: Experiences, evaluations, perspectives. *Med Teach.* 2007;29:219–224.
- 38 Bland CJ, VanLoy W, Wersal L. Lessons learned from a distance-based consulting program to assist faculty development projects. *Acad Med.* 2001;76:776–790.
- 39 Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: Do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *JAMA.* 1999;282:867–874.
- 40 O'Sullivan PS, Irby DM. Reframing research on faculty development. *Acad Med.* 2011;86:421–428.
- 41 Reeves S. Ideas for the development of the interprofessional field. *J Interprof Care.* 2010;24:217–219.

## Appendix 1

### Characteristics of 22 Studies of Faculty Development Programs, Identified in a Systematic Review of the Literature Published Between January 1989 and December 2010

First author, year	Location	Format	Scope*	Study design	Data collection method	Source	Outcomes studied <sup>†</sup>						
							1	2a	2b	3	4a	4b	4c
Alford, 2008 <sup>17</sup>	United States	Workshop	National, individuals	Pre and follow-up (3 months), with comparison group	Surveys	Self-reported	—	X	—	X	—	—	—
Berbano, 2006 <sup>18</sup>	United States	Series/longitudinal	National, individuals	Pre and post	Videos of participants with standardized learners	Observed by others	—	—	—	X	—	—	—
Bland, 2001 <sup>38</sup>	Canada, United States, Puerto Rico	Series/longitudinal	International, groups	During and post	Observations, surveys, progress reports	Self-reported, observed by others	X	—	—	X	X	—	—
Branch, 2009 <sup>19</sup>	United States	Series/longitudinal	National, individuals	Pre, during, and post, with comparison group	Surveys, participant narratives	Self-reported, observed by others	—	X	X	X	X	X	—
Cole, 2004 <sup>20</sup>	United States	Series/longitudinal	National, individuals	Pre and post, with comparison group	Surveys	Self-reported	X	X	X	—	—	—	—
Gozu, 2008 <sup>21</sup>	United States	Series/longitudinal	Local, individuals	Post and follow-up (6–13 years), with comparison group	Surveys	Self-reported	—	X	X	X	—	—	—
Herrmann, 2007 <sup>37</sup>	Germany	Series/longitudinal	National, individuals	Post and follow-up (2–3 months, 24 months)	Interviews, focus groups, surveys	Self-reported	X	X	X	X	—	—	X
Houston, 2004 <sup>22</sup>	United States	Short course	National, groups	Longitudinal (pre, post, 6, 12, 18, 24 months)	Surveys	Self-reported	—	—	—	X	X	—	—
McGaghie, 1990 <sup>23</sup>	United States	Fellowship	Local, individuals	Post, with comparison group	Surveys	Self-reported	—	X	—	X	—	—	—
Morzinski, 2003 <sup>24</sup>	United States	Series/longitudinal	Local, individuals	Longitudinal (1 year pre, during, post, 1-year follow-up)	Surveys, curriculum vitae, progress reports	Self-reported	X	—	X	X	X	—	—

(Appendix continues)

## Appendix 1, Continued

First author, year	Location	Format	Scope*	Study design	Data collection method	Source	Outcomes studied†						
							1	2a	2b	3	4a	4b	4c
Moses, 2009 <sup>25</sup>	United States	Series/longitudinal	Local, individuals	Pre, post, and follow-up (2 years)	Interviews, curriculum vitae	Self-reported	—	X	X	X	—	—	—
Notzer, 2008 <sup>35</sup>	Israel	Workshop	Local, individuals	Pre and follow-up (1 year), with comparison group	Surveys	Observed by others	—	—	X	X	—	—	—
Pandachuck, 2004 <sup>32</sup>	Canada	Short course	Local, individuals	Longitudinal (up to 5 time points), with comparison group	Surveys	Self-reported, observed by others	X	—	X	X	—	—	—
Regan-Smith, 2007 <sup>26</sup>	United States	Other	Local, individuals	During and post, with comparison group	Observations, teaching scores	Observed by others	—	X	X	X	X	—	—
Shields, 2007 <sup>27</sup>	United States	Workshop	Local, individuals	Longitudinal (2 years pre, 3 years post), with comparison group	Surveys, student marks	Observed by others	—	—	X	X	—	X	—
Simpson, 2004 <sup>28</sup>	United States	Series/longitudinal	National, individuals	Retro-pre/post and follow-up (after participants deliver workshop)	Surveys, cost logs, participant goals	Self-reported	X	X	X	X	X	—	—
Steinert, 2005 <sup>33</sup>	Canada	Workshop	Local, individuals	Post and follow-up (18 months)	Surveys	Self-reported	X	X	X	X	X	—	—
Steinert, 2008 <sup>34</sup>	Canada	Other	Local, individuals	Longitudinal (5 years pre, pre, post, 1- and 5-year follow-up)	Observations, surveys, curriculum vitae	Self-reported	X	—	—	X	—	—	—
Stratos, 2006 <sup>29</sup>	United States	Series/longitudinal	National, individuals	Longitudinal (retro-pre/post and before and after participants deliver workshops; at least 2 workshops in first year)	Surveys	Self-reported, observed by others	—	—	X	X	X	X	—
Sullivan, 2005 <sup>30</sup>	United States	Series/longitudinal	National, individuals	Pre, during, and post	Surveys	Self-reported	X	X	X	X	—	—	—
Sullivan, 2006 <sup>31</sup>	United States	Series/longitudinal	National, individuals	Pre, retro-pre/post, and follow-up (6, 12, or 18 months)	Surveys	Self-reported	—	X	X	X	X	—	X
Weurlander, 2008 <sup>36</sup>	Sweden	Series/longitudinal	Local, individuals	Follow-up (1 year)	Interviews	Self-reported	—	X	X	X	—	—	—

\* Scope includes geographic scope of faculty development program (local, national, international) and participant scope (intended for individuals or groups).

† Outcomes studied by source of data (self-reported or observed by others) and type (1: learner reaction; 2a: modification of attitudes/perceptions; 2b: acquisition of knowledge/skills; 3: behavioral change; 4a: changes in organizational practice; 4b: benefits to students/residents; 4c: benefits to patients/communities), using a modified model of Kirkpatrick's educational outcomes.<sup>14</sup>