The BEME Collaboration
Best **E**vidence **M**edical **E**ducation

Faculty development initiatives designed to promote leadership in medical education. A BEME systematic review: BEME Guide No. 19

Running Head: Faculty development for leadership in medical education

Yvonne Steinert, Ph.D.
McGill University, Montreal, Canada

Laura Naismith, M.Phil.
McGill University, Montreal, Canada

Karen Mann, Ph.D.
Dalhousie University, Halifax, Canada

Request for reprints should be addressed to: Yvonne Steinert, Ph.D., Centre for Medical Education, Faculty of Medicine, McGill University, Lady Meredith House, 1110 Pine Avenue West, Montreal, Quebec, H3A 1A3. Tel: (514) 398-2698; Fax: (514) 398-6649; E-Mail: yvonne.steinert@mcgill.ca

I. INTRODUCTION

Successful health care for the 21ˢᵗ century calls for diversification of leadership capabilities and management styles that will enrich our abilities to respond to the needs of all groups.

(Richman et al., 2001, p. 271)

Faculty development refers to those activities that institutions use to *renew* or *assist* faculty in their roles as teachers, researchers and administrators (Centra, 1978; Sheets & Schwenk, 1990). With the increasing complexity of medical education and practice, and recognition of the fact that physicians must assume significant leadership roles (Zaher, 1996), we have witnessed an increase in faculty development activities designed to enhance leadership in medical education. These include formal training programs (Morahan *et al*., 1998) and fellowships (Korschun *et al*., 2007); workshops (Steinert *et al.*, 2003a) and seminars (Woods & Griggs, 1994); faculty internships (Seavey & Hiller, 1984); and mentoring programs (Garman *et al.*, 2001). Some faculty development activities have also targeted organizational systems and development (Aluise *et al.*, 1985), whereas others have focused on academic and career skills (Morzinski & Simpson, 2003).

To date, only a few publications have reviewed faculty development activities that focus on leadership. For example, Bogdewic *et al.* (1997) reviewed the curricula of several national faculty development fellowship programs to identify major emphases, strategies and outcomes regarding organizational and leadership development for academic physicians. They identified three types of organizational and leadership development strategies: isolated faculty development workshops at national meetings, longitudinal faculty development fellowship programs, and organizational change efforts within an academic department or residency program. Indicators of success included career satisfaction, retention, and attainment of higher academic rank among those who participated. More recently, Gruppen *et al.* (2006) reviewed a series of educational fellowship programs that produced leaders in medical education. Outcomes included academic promotions, new leadership positions, and scholarly productivity. Importantly, neither review was systematic nor comprehensive; in addition, there was no comparison of different faculty development interventions, and little attention was paid to the impact of faculty development on the organizations in which individuals work.

The goal of this article was to systematically review the literature on faculty development programs that target leadership capabilities among faculty members and try to assess which programs (e.g. workshops; training courses; fellowships) were most effective. In addition, we wanted to identify key outcomes (e.g. change in learning; change in behaviour; change in the system) as well as program "features" associated with effectiveness in order to make recommendations for practice and research.

Best Evidence Medical Education

The Best Evidence Medical Education (BEME) Collaboration (www.bemecollaboration.org/) is an international group of individuals, universities and organizations committed to synthesizing and disseminating the latest educational research findings in order to provide a basis for informed decision-making (Harden *et al.,* 1999).

In 2001, the BEME Collaboration established an international Faculty Development Topic Review Group (TRG) to review the "best evidence" in faculty development designed to improve teaching effectiveness in medical education. This report, published in 2006 (Steinert *et al.*, 2006), highlighted the following results: high satisfaction with faculty development programs; positive changes in attitudes toward faculty development and teaching; increased knowledge of educational principles, and gains in teaching skills; self-reported behaviour changes as well as observed changes in performance; and few reported changes in organizational practice or student learning.

In 2008, the lead investigators of the first BEME review (YS and KM) decided to examine the faculty development literature as it pertains to the development of leadership among faculty members. The current report describes the review process and synthesizes the results under the following headings:

- ▪ **OBJECTIVES** – which summarizes the overall objectives of this review
- ▪ **REVIEW QUESTION** – which describes the primary review question

- **REVIEW METHODOLOGY** – which includes group formation, the pilot process, analysis of outcomes, inclusion/exclusion criteria, search strategy and sources of articles, and selection methods and judgment of methodological quality
- **DATA MANAGEMENT TECHNIQUES** – which includes data extraction, analysis and synthesis
- **REVIEW FINDINGS** – which describes an overview of the articles included in this review, narrative comments on both the review findings and the methodological quality of the studies, and a summary of the results, by program type
- **DISCUSSION** – which highlights the major findings of this review by summarizing outcomes, describing "key features" of faculty development activities in this area, identifying avenues for future development as well as methodological issues, and discussing the strengths and limitations of this review
- **CONCLUSION** – which describes implications for future practice and research based on the review's findings

II. OBJECTIVES

The objectives of this review were to determine the effects of faculty development activities on faculty members' leadership capabilities and to assess the impact of these activities on the institutions in which these individuals work. To achieve these objectives, we concentrated on three main types of faculty development programs: (1) those that focused primarily on leadership; (2) those that addressed leadership as part of a broader focus on educational development; and (3) those that addressed leadership as part of a broader focus on academic career development. Moreover, based on our previous article (Steinert *et al.,* 2006), we limited this review to faculty development programs intended for faculty members in medicine; we did not examine programs specifically designed for residents or other health care professionals (e.g. nurses; dentists). However, all types of faculty development interventions (e.g. workshops, short courses and seminars, and fellowships) were included in this review.

III. REVIEW QUESTION

The primary research question for this topic review was as follows:

> *What are the effects of faculty development interventions designed to promote leadership on the knowledge, attitudes, and skills of faculty members in medicine, and on the institutions in which they work?*

In addition, we tried to answer the following questions, based on lessons learned in our previous review:

- What characterizes the faculty development activities that have been described – and what are the "features" that make them effective?
- What are the methodological strengths and weaknesses of the reported studies?
- What are the implications of this review for faculty development practices and ongoing research in this area?

IV. REVIEW METHODOLOGY

Group Formation

A national Topic Review Group (TRG) of three individuals was constituted to conduct this review. Based on our previous experience with an international group and the challenge of face-to-face meetings with individuals on three continents, we chose to keep the TRG small. The two lead investigators had extensive experience in faculty development and medical education as well as expertise in educational research methodology; the addition of a research associate was invaluable in moving the task forward.

The Pilot Process

The pilot process consisted of two phases:

Phase I

To initiate the process, the lead reviewer conducted a scoping search and adapted the original Faculty Development BEME Coding Sheet to include a focus on leadership development. Two of the TRG members (YS and LN) as well as an external associate then reviewed five articles to determine the scope of the review, to refine the review question, and to assess the utility of the revised coding sheet. As a result, we modified the BEME Coding Sheet by: (1) adding a leadership category to *Expected Learning Outcomes of the Intervention* and consolidating the

teaching outcomes into one category; (2) including a section to capture the intervention's *Definition/Operationalization of Leadership*; (3) expanding the section on demographic characteristics and adding leadership roles (at entry to the program) and selection procedures to *Context (Target Population)*; and (4) adding program title, content, and leadership focus to the *Stated Intervention*.

This phase also led to the decision that we would address three types of faculty development interventions: those with a primary focus on leadership; those that included leadership as a component of a larger, more comprehensive faculty development program focusing on educational development; and those that included leadership as a component of a larger program focusing on academic career development.

Phase II

The second phase consisted of a pilot review of an additional five articles by two TRG members (YS and LN). This process helped us to finalize the BEME Coding Sheet (which is available as **Appendix 1** on the BEME website) and to determine the full scope of the literature search, a process for working together, and additional needs for reviewer training.

Analysis of Outcomes

We used Kirkpatrick's model of educational outcomes (1994) to classify and analyze outcomes (see **Figure 1**). The model describes four levels of outcome: learners' reaction (to the educational experience); learning (which refers to changes in attitudes, knowledge and skills); behaviour (which refers to changes in practice and the application of learning to practice); and results (which refers to change at the level of the organization). In line with our previous review (Steinert *et al*., 2006), we used Freeth *et al*.'s adaptation (2002) of the Kirkpatrick model, which divided learning into two categories: modification of attitudes/perceptions and acquisition of knowledge and skills. We also further divided behaviour into two separate categories: *self-reported* changes in practice and *observed* changes in practice, including new leadership positions. Kirkpatrick's model (1994), which is not meant to be hierarchical, has been used by other BEME groups (e.g. Freeth *et al*., 2003; Issenberg *et al.*, 2005).

Inclusion/Exclusion Criteria

On the basis of the pilot process, the following criteria were used to select articles for review:

Leadership Focus – Training interventions that included the development of leadership (knowledge, skills and/or attitudes) as an expected learning outcome were selected for this review. In the initial stages, leadership was broadly envisioned to include all aspects of management and administration.

Empirical Focus – Articles that described training interventions that had been implemented and evaluated were selected. All types of faculty development interventions, regardless of duration, were included. Articles that solely provided descriptions of interventions, conceptual frameworks and/or recommendations for training interventions were excluded.

Target Population – Faculty development interventions for basic science and clinical faculty members were selected. Interventions designed to improve leadership among residents-in-training or other health care professionals (e.g. nurses) were excluded. However, interprofessional faculty development activities that included faculty members in medicine were included.

Study Design – Consistent with our previous review, we included all study designs across the positivist (empirical observation and measurement), interpretist (construction of understanding), and participatory (action research) paradigms (Creswell, 2003; Freeth *et al.*, 2005). However, only studies that included outcome data beyond participant satisfaction were examined. While participant satisfaction is important, we wanted to explore evidence of learning and change. With the relatively small number of research articles in this field, and consistent with other BEME reviews, we elected not to restrict our search to the "gold standard" of randomized controlled trials as is common in systematic reviews in medicine (Egger *et al.*, 2001).

Year of Publication – We searched for articles that were published between 1980 and 2009. Based on expert recommendations, we also reviewed two articles that had been accepted for

publication and were to appear in 2010. We chose 1980 based on our experience with our previous review.

Language and Geography – While we did not make any exclusions on the basis of language or geography, only English-language results appeared in our database searches and we did not make further efforts to hand search articles published in other languages.

Search Strategy and Sources of Articles

A thorough search involving multiple approaches was used to reduce bias in the review process (Centre for Reviews and Dissemination, 2009). Due to the interdisciplinary nature of this topic, we conducted our literature search using six databases, representing medicine (Medline, EMBASE, CINAHL, and Web of Science), education (ERIC) and management (ABI/Inform). Keywords used for the inclusion criteria included the following:

- Leadership Focus - leadership, management, administration, executive, change agent
- Empirical Focus - faculty development, in-service training
- Target Population - doctor, medic, physician, faculty
- Study Design - evaluate, assess, impact, outcome

The database searches were conducted with the assistance of a McGill health sciences librarian. A copy of the search strategy is included in **Appendix II** (which is available on the BEME website).

The medical education literature is known to be fragmented and poorly indexed, particularly with respect to mixed methods research (Maudsley, 2011). Therefore, it was important for us to pursue alternate approaches to finding relevant literature. To this end, we conducted hand searches of personal files and reference sections of all retrieved articles. We also solicited expert recommendations from three prominent scholars in the field and wrote to six authors of previously identified articles for additional evaluation reports and suggestions.

Selection Methods and Judgment of Methodological Quality

The database search yielded a total of 530 unique records. A two-stage process, outlined in **Figure 2**, was employed to select studies eligible for review. Initially, each title and abstract was

evaluated by one of the TRG members (LN) and an external associate to ascertain whether the article related to faculty development and the development of leadership capabilities among faculty members. This resulted in 46 (8.7%) articles. Discrepancies in judgment between the two reviewers were resolved through discussion. The hand search (of all reference lists and the lead reviewer's own files) resulted in an additional 24 articles related to leadership development; expert recommendations yielded five articles. Full texts of all 75 articles that related to leadership development were retrieved for further examination.

For the second step, two individuals (LN and YS) reviewed the full text of each retrieved article to apply the inclusion criteria. Forty-eight articles related to leadership development and included outcome data beyond satisfaction ratings. Thirty-one (65%) of these articles came from the database search, 12 (25%) came from the hand search and 5 (10%) came from expert recommendations. The remaining articles described faculty development programs with no evaluation data or consisted of conceptual approaches to professional development; they were all eliminated.

V. DATA MANAGEMENT TECHNIQUES

Data Extraction, Analysis and Synthesis

The modified BEME coding sheet was used to facilitate data extraction. Data were collected on the following items:

- Expected learning outcomes of the intervention
- Definition/operationalization of leadership and/or conceptual framework used
- Characteristics of the target population
- Aim/goal of the study
- Description and impact of the intervention
- Evaluation methods, including study design, data collection methods and data sources
- Study quality and strength of findings
- Avenues for further research
- New insights and implications for faculty development

Members of the TRG reviewed and coded each set of articles in pairs. All coding sheets were then compiled into a Microsoft Excel spreadsheet. As in our previous review (Steinert *et al.*, 2006), this database provided a means of visually comparing the information and facilitated frequency counts of important categories. The spreadsheet was then returned to one reviewer per team who was asked to resolve coding differences. Where necessary, the lead reviewer assisted in resolving differences; she also read all of the articles and coding sheets to ensure uniformity in approach.

As the search yielded both quantitative and qualitative studies, we used an integrated design to synthesize findings (Sandelowski *et al.*, 2006). An integrated design assumes that both quantitative and qualitative studies can address the same research question and that results from both types of studies can be synthesized. For example, learning may be measured quantitatively using a self-report questionnaire (e.g. McDade *et al.*, 2004) or qualitatively based on a thematic analysis of project content (e.g. McCurdy *et al.*, 2004). We interpreted these results as extending and confirming each other, rather than reflecting different phenomena (Sandelowski *et al.*, 2006).

VI. REVIEW FINDINGS

This review was based on 48 articles, describing 41 studies of 35 interventions, all of which focused on faculty development to promote leadership. This section is organized into three main components:

(1) Those studies in which leadership was the primary focus of the intervention;

(2) Those studies in which leadership was a component of a broader focus on educational development; and

(3) Those studies in which leadership was a component of a broader focus on academic career development.

Tables 1, 2 and **3** summarize the articles that were reviewed.

CATEGORY 1: LEADERSHIP AS THE PRIMARY FOCUS OF THE INTERVENTION

Description of the Articles

We retrieved 19 articles in which leadership was the primary focus of the intervention. The publication dates of these articles ranged from 1985 to 2010, with the majority of articles (14, 72%) published in 2001 or later. Six of these studies described the same intervention, the Executive Leadership in Academic Medicine (ELAM) fellowship program (Dannels *et al.*, 2009; Dannels *et al.*, 2008; McDade *et al.*, 2004; Morahan *et al.*, 2010; Richman *et al.*, 2001; Sloma-Williams *et al.*, 2009). Thus, 19 studies represent 14 interventions. To simplify the presentation of our findings, the following section will report on the 14 interventions described in the literature. However, all studies were reviewed separately for methodological quality.

Description of the Interventions

Setting

Most of the 14 leadership interventions took place within the context of medical schools or academic health centres in the United States (12, 86%). One intervention (7%) took place in Canada (Steinert *et al.*, 2003a) and another in the United Kingdom (Fox *et al.*, 2001). Eight (57%) of the programs were considered "local" in nature (Bachrach, 1997; Duda, 2008; Korschun *et al.*., 2007; McCurdy *et al.*, 2004; Morahan *et al.*, 1998; Steinert *et al.*, 2003a; Stoller *et al.*, 2007; Woods & Griggs, 1994), in that they took place at the participants' home institutions, while six (43%) were regional and/or national programs that welcomed individuals from different schools and institutions (Aluise *et al.*, 1985; Coleman *et al.*, 1998; Fox *et al.*, 2001; Leslie *et al.*, 2005; Morahan *et al.*, 2010; Osborn & DeWitt, 2004).

Program Participants

The majority of leadership interventions targeted clinical faculty members, primarily in family medicine (Aluise *et al.*, 1985; Steinert *et al.*, 2003a) and pediatrics (Leslie *et al.*, 2005; Osborn & DeWitt, 2004). Many interventions did not report demographic data, and only one program (ELAM) targeted women specifically. Two interventions (14%) targeted junior faculty (Duda, 2008; Leslie *et al.*, 2005) and three (21%) targeted senior faculty (Aluise *et al.*, 1985; Bachrach,

1997; Morahan *et al.*, 2010). Nine interventions (64%) included medical professionals exclusively, while five (36%) were interprofessional in nature and included dentists (McCurdy *et al.*, 2004; Morahan *et al.*, 2010), nurses (Korschun *et al.*, 2007; McCurdy *et al.*, 2004), pharmacists (McCurdy *et al.*, 2004), public health professionals (Korschun *et al.*, 2007; Morahan *et al.*, 2010), and professional administrative staff (Bachrach, 1997; Korschun *et al.*, 2007; Morahan *et al.*, 1998).

Nomination of participants was the most popular selection procedure (5, 36%), with the next most popular procedures being by application (3, 21%) and open registration (3, 21%). Two interventions (14%) actively recruited participants (Aluise *et al.*, 1985; Fox *et al.*, 2001), and in one intervention, participants were selected based on diversity of specialty, practice venue, gender, race, and age (Stoller *et al.*, 2007). Based on the information provided, it did not appear that participation was compulsory in any intervention. The number of participants per intervention ranged from 7 to 152, with a mean of 36 participants per cohort. The smallest number of participants in an evaluation study was 7 (Woods & Griggs, 1994) and the largest was 569 (longitudinal evaluation of the ELAM program, Morahan *et al.*, 2010).

Focus of the Intervention
All interventions in this category focused explicitly on leadership development. Eleven interventions (79%) took a broad view of leadership and seven articulated specific program objectives, such as the following: to increase faculty members' leadership skills and stimulate changes in the institutional culture (Osborn & DeWitt, 2004); to increase personal awareness and develop an appreciation for a leader's role in a complex organization (Morahan *et al.*, 1998); and to foster participants' conceptual understanding of leadership (Leslie *et al.*, 2005). Most interventions addressed multiple topics. Frequently mentioned aspects of leadership included: conflict management and negotiation (9), budgeting and financial management (7), leadership theory and concepts (6), people management and performance issues (6), networking, team-building and mentoring (6), organizational structure and culture (5), change management (5), strategic planning and problem-solving (4), time management (4), and personal leadership styles (4). Three interventions focused exclusively on one topic: budgeting and financial management

(Woods & Griggs, 1994), continuous quality improvement (Coleman *et al.*, 1998) and change management (Fox *et al.*, 2001).

Program Type and Duration

Interventions were classified according to the authors' terminology. Three interventions (21%) were described as training or development programs, lasting from 6 months to 1 year (Bachrach, 1997; Leslie *et al.*, 2005; Osborn & DeWitt, 2004). Three interventions (21%) were described as part-time fellowships that ranged in duration from 6 months to 1 year (Korschun *et al.*, 2007; Morahan *et al.*, 2010; Morahan *et al.*, 1998). Both fellowships and training programs typically included a combination of periodic group instructional sessions, one-on-one mentoring, and individual or small group project work. Three interventions (21%) were described as courses that ranged in duration from one day to one year (Duda, 2008; Fox *et al.*, 2001; McCurdy *et al.*, 2004); three interventions (21%) were described as workshops lasting from 8 hours to 3.5 days (Aluise *et al.*, 1985; Coleman *et al.*, 1998; Steinert *et al.*, 2003a), and one intervention was described as a 6-hour seminar series (Woods & Griggs, 1994). Stoller *et al.* (2007) originally called their intervention a "program" and then relabeled it a "course" after it was "radically restructured" in 2002 (p. 238). The following sample titles give a flavor of the intervention type: Executive Leadership in Academic Medicine (Dannels *et al.*, 2009; Dannels *et al.*, 2008; McDade *et al.*, 2004; Morahan *et al.*, 2010; Richman *et al.*, 2001; Sloma-Williams *et al.*, 2009): Executive Skills for Medical Faculty (Steinert *et al.*, 2003a); Woodruff Leadership Academy (Korschun *et al.*, 2007); Executive Development Program (Bachrach, 1997); Allegheny Leadership Institute (Morahan *et al.*, 1998); Administrative Colloquium (McCurdy *et al.*, 2004); Faculty Development Scholars, Executive Leadership Track (Osborn & DeWitt, 2004); Young Pediatric Leaders for the 21st Century (Leslie *et al.*, 2005); and Executive Program in Practice Management/Leading in Health Care Course (Stoller *et al.*, 2007).

Instructional Methods

Instructional methods varied widely, with a particular focus on the use of experiential learning to provide structured practice opportunities (9, 69%). Sample practice opportunities included developing a mission statement (McCurdy *et al.*, 2004; Osborn & DeWitt, 2004), personal goal setting (Steinert *et al.*, 2003a), and planning team meetings (Coleman *et al.*, 1998). Five

interventions (38%) used projects, both individual (McCurdy *et al.*, 2004; Morahan *et al.*, 2010) and group (Korschun *et al.*, 2007; Morahan *et al.*, 1998; Stoller *et al.*, 2007), to reinforce face-to-face sessions. Additional instructional methods included small group discussions, case-based learning, role plays, and mentoring. Very few studies used computer-based materials. Fox *et al.* (2001) were notable for delivering their entire course at a distance using the Internet, and Richman *et al.* (2001) used computer simulations to teach budgeting and financial management. The majority of interventions (8, 57%) were either developed in response to a stated need (e.g., McCurdy *et al.*, 2004) or explicitly used needs assessments in the design of their curricula and evaluation measures.

Innovative program components included: writing personal and cohort mission statements (McCurdy *et al.*, 2004; Osborn & DeWitt, 2004); implementing a 360-degree review by direct reports, peers, and supervisors (Korschun *et al.*, 2007); developing a business plan in teams (Morahan *et al.*, 1998; Stoller *et al.*, 2007); interviewing influential leaders at home institutions (Morahan *et al.*, 2010); using a behavioural change contract to implement a specific leadership change (Leslie *et al.*, 2005); implementing a formal graduation ceremony (Bachrach, 1997; McCurdy *et al.*, 2004); and providing the opportunity to become a member in a professional society (Morahan *et al.*, 2010).

Methodological Quality of the Studies

In this section, we shift our focus from describing the 14 interventions in this category to considering the methodological quality of the 19 studies.

Study Goal and Theoretical Framework

Eighteen of the 19 studies (95%) stated their objectives (e.g. to describe, implement and evaluate a faculty development initiative). Some reports described more specific objectives, outlining a particular study question, such as assessing whether participation in the intervention enhanced the leadership and career development of women faculty as compared to women from two comparison groups (Dannels *et al.*, 2008), or assessing the frequency with which business plans submitted by course attendees were implemented (Stoller *et al.*, 2007). In examining the extent to which the studies cited the relevant literature, we classified seven (37%) as doing so

comprehensively, eight (42%) as doing so adequately, and four (21%) as doing so in a limited fashion. Ten studies (53%) placed their work within a conceptual framework, but only eight (42%) provided an explicit definition or operationalization of leadership.

Study Design

The majority of the 19 studies were quantitative in nature (12, 63%) and attempted to measure specific outcomes of participation. Two qualitative studies focused on describing and understanding participants' experiences of the intervention (Bachrach, 1997; Sloma-Williams *et al.*, 2009) and four mixed-methods studies pursued both of these objectives (Fox *et al.*, 2001; Korschun *et al.*, 2007; Morahan *et al.*, 1998; Steinert *et al.*, 2003a). One study (Aluise *et al.*, 1985) used an action research methodology to design, plan, implement, and evaluate a faculty development intervention.

Of the 16 studies that included a quantitative component, 15 employed a quasi-experimental design using a single group with no comparison. Only one study compared participants to two non-equivalent control groups (Dannels *et al.*, 2008). Nine of the 16 studies with a quantitative component used a pre-test/post-test design (three retrospective), while seven used post-test measures only. One study used an immediate post-test only, six used delayed post-tests only, eight used both immediate and delayed post-tests, and one did not specify the timing of the post-test. Of the five studies that included a qualitative component, two were case studies (Bachrach, 1997; Sloma-Williams *et al.*, 2009); the others did not specify a particular study design.

Data Collection Methods, Sources and Analyses

Methods to evaluate these faculty development interventions included end-of-workshop questionnaires, pre- and post-test measures to assess attitudinal or cognitive change, self-assessment of post-training performance, and end-of-program interviews. Fourteen studies (74%) used self-report questionnaires that were designed by the study authors. McDade *et al.* (2004) used expert review to validate their questionnaire, and Cronbach's alpha was reported as a measure of internal consistency in two additional studies (Dannels *et al.*, 2008; Dannels *et al.*, 2009); however, most questionnaires did not appear to have undergone a formal validation process. Effect sizes were only reported for the comparative study (Dannels *et al.*, 2008). Seven

studies (37%) used interviews, but they did not describe how the data were analyzed. Two studies (11%) used documentary sources (McCurdy *et al.*, 2004; Stoller *et al.*, 2007) and one study (5%) was based on a database analysis of program participants (Morahan *et al.*, 2010). Program participants were the primary data source in 17 (89%) studies. Dannels *et al.* (2009) reported on impact as observed by a third party (i.e. medical school deans), Morahan *et al.* (1998) interviewed organizational executives as well as program participants, and two studies (Morahan *et al.*, 1998; Stoller *et al.*, 2007) reviewed participants' business plans.

Study Quality and Strength of Findings

Study quality was rated on a five-point scale (1 = low; 5 = high), and reviewers were asked to indicate study strengths and weaknesses. Based on our previous review (Steinert *et al.*, 2006), we chose to use one overall rating for this variable rather than sub-scales. Strength of findings was rated on a five-point scale with specific anchors (1 = no clear conclusions can be drawn; 3 = conclusions can probably be based on results; 5 = results are unequivocal). Across raters, the mean rating for study quality in this section was 2.8 and the mean rating for strength of findings was 3.0.

Reported Outcomes

For each intervention that focused primarily on leadership, outcomes were assessed at multiple levels. In total, nine interventions (64%) reported results at the ***reaction*** level, which included participant satisfaction, perception of program usefulness, and value of the activity. Thirteen interventions (93%) assessed ***learning***, which included changes in attitudes (8, 57%) as well as changes in knowledge or skills (13, 93%). Eleven interventions (79%) indicated ***self-reported behaviour*** changes, while five interventions (36%) reported ***observed behaviour*** changes. Four interventions (29%) reported outcomes at the ***results*** level in terms of changes in organizational practice.

Level 1 – REACTION: Satisfaction was usually measured on a Likert scale or a comparable categorical scale, with ratings from poor to excellent. Most of the participants were positive about their experiences and rated the faculty development course content (Leslie *et al.*, 2005) and overall experience (Aluise *et al.*, 1985) highly. Many participants reported that the interventions

were useful (Coleman *et al.*, 1998; McDade *et al.*, 2004; Steinert *et al.*, 2003a) and of both personal and professional benefit (Duda, 2008; Morahan *et al.*, 1998). They valued the interactive, experiential and collegial nature of the programs offered as well as the direct applications to personal settings, and they recommended the continuation of the intervention (Aluise *et al.*, 1985; Woods & Griggs, 1994). Morahan *et al.* (1998) reported that the strongest impact of the intervention, as identified through participant self-report, was on the relationships the participants built with others in the program. In contrast, Fox *et al.* (2001) reported that participants appreciated the flexibility provided by the Internet-based distance learning course, although they often viewed the interaction with other participants and the course facilitator as inadequate.

Level 2a – LEARNING: Outcomes at this level addressed attitudes, which were measured using both self-report questionnaires and interviews. Attitudinal changes included a broader organizational perspective (Bachrach, 1997) and increased commitment to institutional vision (Bachrach, 1997; Korschun *et al.*, 2007). Participants also reported feeling more confident and self-aware (Aluise *et al.*, 1985; Sloma-Williams *et al.*, 2009) and more motivated to consider pursuing a leadership position in the future (Duda, 2008; Korschun *et al.*, 2007). In addition, participants reported both intent to change specific leadership behaviours (e.g. manage time differently, be more group-centred as a leader) (Steinert *et al.*, 2003a) and increased self-efficacy in using existing abilities towards specific leadership goals, such as chairing a task force on recommendations for implementing computer-assisted instruction in the medical school curriculum (Sloma-Williams *et al.*, 2009). While the majority of attitudinal changes were self-reported, in one study (Dannels *et al.*, 2009) the deans reported that the ELAM intervention increased participants' confidence and broadened their perspectives about academic medicine.

Level 2b – LEARNING: Changes in knowledge included increased understanding of targeted curriculum content areas, including leadership theories and constructs, leadership styles and skills, strategic planning, financial management, and organizational development and change management (Coleman *et al.*, 1998; Dannels *et al.*, 2008, 2009; Fox *et al.*, 2001; Leslie *et al.*, 2005; McCurdy *et al.*, 2004; McDade *et al.*, 2004; Morahan *et al.*, 1998; Osborn & DeWitt, 2004; Woods & Griggs, 1994). In some studies, these changes were demonstrated through pre-

post self-report measures on knowledge of particular curricular areas (Fox *et al.*, 2001; McCurdy *et al.*, 2004; Richman *et al.*, 2001; Woods & Griggs, 1994) and pre-post measures of change in leadership skills (Dannels *et al.*, 2008; Leslie *et al.*, 2005; McDade *et al.*, 2004; Morahan *et al.*, 1998; Osborn & DeWitt, 2004; Richman *et al.*, 2001). In other studies, post measures only were used (Coleman *et al.*, 1998; Steinert *et al.*, 2003a). Qualitative interview data showed that participants gained a clearer understanding of their own roles and responsibilities and recognition of their own management styles, including strengths and weaknesses (Bachrach, 1997). Participants in the Sloma-Williams *et al.* (2009) study referred to specific increases in financial management and communication skills, a general sense of overall skill development, and an improved ability to perform under stress. As with attitudinal changes, most changes in knowledge were self-reported. However, Dannels *et al.* (2009) found that deans reported enhanced leadership, business and management skills among their ELAM participants and half of the participants in the Sloma-Williams *et al.* (2009) study reported either external validation of their leadership abilities or the acquisition of new skills.

Level 3a – BEHAVIOUR (self-reported): Self-reported behavioural changes were identified through both questionnaires and interviews. These included applying specific techniques learned in the intervention to the workplace, such as reorganizing a department and broadening representation on an executive committee (Aluise *et al.*, 1985), developing a budget for a small grant or project and successfully negotiating for departmental resources (Woods & Griggs, 1994), planning structured team meetings (Coleman *et al.*, 1998), and determining short-term goals and handling paper more effectively (Steinert *et al.*, 2003a). Many participants assumed new leadership roles and responsibilities following the intervention, including becoming presidents of national organizations, department chairs, deans, vice presidents, provosts or chief executive officers (Dannels *et al.*, 2008; Duda, 2008; Korschun *et al.*, 2007; Morahan *et al.*, 2010; Osborn & DeWitt, 2004), and they established collaborations based on networking during the intervention (Korschun *et al.*, 2007; Leslie *et al.*, 2005). Participants also reported pursuing ongoing leadership training as a result of the intervention (Aluise *et al.*, 1985; Dannels *et al.*, 2008; Duda, 2008; Leslie *et al.*, 2005), achieving individual leadership goals (Leslie *et al.*, 2005), and achieving academic promotion (Dannels *et al.*, 2008; Korschun *et al.*, 2007; Osborn & DeWitt, 2004). In the comparative study, Dannels *et al.* (2008) noted that ELAM graduates

were more likely to be departmental or committee chairs than were women of a similar profile who did not participate in the program. There were no significant differences in the proportion of women in either group holding national leadership positions.

Level 3b – BEHAVIOUR (observed): Few studies assessed observed behavioural changes. Two studies (Morahan, 1998; Stoller *et al.*, 2007) reported a high rate of implementation of participants' business plans (67% and 61%, respectively), though neither specified how this was measured. Based on an analysis of their program database, Morahan *et al.* (2010) reported that 35 of 569 alumni of the ELAM program held high-level leadership positions (vice president, provost, president), both within and outside academia, and that 25% of current deans of medical, dental, and public health schools were ELAM graduates. Bachrach (1997) noted that 93% of program participants were retained at the institution three years later, with the other 7% holding high profile academic leadership positions at other institutions. Dannels *et al.* (2009) reported that ELAM fellows made a positive impact on their schools and that the deans who had sponsored three or more fellows responded significantly more positively to the evaluation questionnaire than those who sponsored fewer.

Level 4a – RESULTS: Very few studies examined change at the organizational/systems level. One study identified self-reported changes in organizational practices, including a shift to mission-based budgeting, an increased emphasis on educational scholarship in promotion and tenure, and increased collaboration between community and academic faculty (Osborn & DeWitt, 2004). Another study reported increased inter- and intra-departmental collaboration (Bachrach, 1997), while a third reported that the intervention helped to implement organizational re-structuring (Morahan *et al.,* 1998). According to Morahan *et al*. (2010), the ELAM program model was applied to the development of two independent leadership programs.

CATEGORY 2: LEADERSHIP AS A COMPONENT OF AN EDUCATIONAL DEVELOPMENT PROGRAM

Description of the Articles

We retrieved 21 articles that focused on the development of leadership within the context of improving teaching effectiveness and/or enhancing educational excellence. The publication dates

of these articles ranged from 2002 to 2010, with the majority of articles (13, 62%) published in 2006 or later. Eight of the 21 articles were retrieved from a special issue of *Academic Medicine* (volume 81:11) that focused on medical education fellowships (Frohna *et al.*, 2006; Muller & Irby, 2006; Robins *et al.*, 2006; Rosenbaum *et al.*, 2006; Searle *et al.*, 2006; Simpson *et al.*, 2006; Steinert & McLeod, 2006; Wilkerson *et al.*, 2006). As these reports included both program descriptions and evaluative data, they met the inclusion criteria of this review. In some cases, however, the evaluative data were a summary of the findings of other research studies. For example, Frohna *et al.* (2006) summarized the findings of Gruppen *et al.* (2003). In such cases, only the original research study was assessed for methodological quality, though results are reported from both articles where appropriate. There were other cases where articles described the same data set. For example, Burdick *et al.* (2010) extended the findings of Burdick *et al.* (2006) and Norcini *et al.* (2005). In these cases, the most comprehensive study was assessed for methodological quality, though results are reported from all studies where appropriate. In total, therefore, the 21 articles represent 14 studies that describe 14 interventions.

Description of the Interventions

Setting

Twelve of the 14 educational development interventions (86%) took place in the United States, one (7%) took place in Canada (Steinert & McLeod, 2006), and another was an international fellowship for faculty in developing countries in Latin America, Africa, and Asia (Burdick *et al.*, 2010). This intervention included a short residency period in the United States, followed by distance learning sessions in the participants' home countries. Compared to the previous section in which interventions focused primarily on leadership, a much larger proportion of these faculty development interventions (10, 71%) were "local" in nature. Three interventions were national programs (Peters *et al.*, 2002; Simpson *et al.*, 2004; Sullivan *et al.*, 2009) and one, as described above, was international in nature (Burdick *et al.*, 2010).

Program Participants

The majority of these interventions targeted clinical faculty members across specialties, including primary care (Simpson *et al.*, 2006) and pediatrics (Simpson *et al.*, 2004). Many did not report demographic data. In comparison to the interventions described in the previous

section, none of the educational development interventions specifically targeted women, junior faculty or senior faculty. Six interventions (43%) were targeted exclusively at the medical profession, while eight (57%) were interprofessional in nature and included dentists (Gruppen *et al.*, 2003; Robins *et al.*, 2006; Steinert & McLeod, 2006), pharmacists (Robins *et al.*, 2006), nurses (Burdick *et al.*, 2010; Gruppen *et al.*, 2003; Robins *et al.*, 2006; Steinert & McLeod, 2006; Sullivan *et al.*, 2006, 2009), public health professionals (Gruppen *et al.*, 2003), professional administrative staff (Muller & Irby, 2006), veterinary scientists (Srinivasan *et al.*, 2007), and other health care professionals (Steinert & McLeod, 2006). The majority of participants in all interventions were medical professionals. The most popular selection procedure was by application (7, 50%), followed by nomination of participants (4, 29%). Other selection methods included self-nomination (1, 8%) and open registration (1, 8%). Notably, six interventions (43%) required prospective participants to demonstrate that they had the explicit support of their department chairs. Based on the information provided, it did not appear that participation was compulsory in any intervention. The number of participants per intervention ranged from 5 to 107, with a mean of 23 participants per cohort. Most evaluation studies reported results from multiple cohorts. The smallest number of participants in an evaluation study was 15 (Steinert *et al.*, 2003b) and the largest was 114 (Sullivan *et al.*, 2006).

Focus of the Intervention

The objectives of the interventions were multi-faceted. Specific leadership objectives included: the development of skilled educational leaders (Burdick *et al.*, 2010; Gruppen *et al.*, 2003; Muller & Irby, 2006; Srinivasan *et al.*, 2007); enhancing attitudes as educational leaders (Searle *et al.*, 2006); preparing faculty for local and national leadership roles (Robins *et al.*, 2006; Steinert & McLeod, 2006; Wilkerson *et al.*, 2006); promoting educational leadership in the context of providing peer training (Rosenbaum *et al.*, 2005); and creating leadership to support education in a particular specialty, such as pediatrics (Simpson *et al.*, 2004) and palliative care (Sullivan *et al.*, 2006, 2009). All interventions addressed the development or promotion of teaching and learning abilities; in addition, eight (57%) focused on developing research and scholarship abilities, eight (57%) addressed career development issues such as promoting professional development and strengthening curricula vitae (CVs) for promotion and tenure purposes, and one (7%) focused on clinical care.

Eight interventions addressed leadership explicitly within their curricula, by including topics such as organizational change management (4), leadership styles (3), running effective meetings (2), team-building and group decision processes (2), conflict resolution (1), qualities of effective leaders (1), and handling difficult conversations (1). One additional intervention addressed leadership skills in the context of training faculty to teach leadership-related topics such as quality improvement and cost-effectiveness (Peters *et al.*, 2002).

Program Type and Duration

Interventions were classified according to the authors' terminology. Eight interventions (57%) were described as scholars programs, lasting from six months to three years. Five of these interventions were specifically referred to as Teaching Scholars Programs (Muller & Irby, 2006; Robins *et al.*, 2006; Rosenbaum *et al.*, 2005; Srinivasan *et al.*, 2007; Steinert & McLeod, 2006). Three interventions (21%) were described as fellowships, all lasting two years part-time (Burdick *et al.*, 2010; Searle *et al.*, 2006; Wilkerson *et al.*, 2006). Two interventions were described as faculty development programs (Peters *et al.*, 2002; Sullivan *et al.*, 2006), lasting six months, and another consisted of a series of discrete modules, each lasting four or five months (Simpson *et al.*, 2006). All interventions were implemented to allow for part-time involvement of participating faculty. Sample titles include: Teaching Scholars Program (Muller & Irby, 2006; Robins *et al.*, 2006; Rosenbaum *et al.*, 2005; Srinivasan *et al.*, 2007; Steinert & McLeod, 2006); Medical Education Scholars Program (Gruppen *et al.*, 2003); Foundation for Advancement of International Medical Education and Research Fellowship (Burdick *et al.*, 2010); Medical Education Fellowship (Wilkerson *et al.*, 2006); Education Scholars Fellowship Program (Searle *et al.*, 2006); Excellence in Clinical Education and Leadership (Simpson *et al.*, 2006); Faculty Development Scholars, Educational Scholarship Track (Simpson *et al.*, 2004); and Program in Palliative Care Education and Practice (Sullivan *et al.*, 2006).

Instructional Methods

As observed in the studies which focused primarily on leadership development, most educational development interventions utilized a variety of instructional methods. Projects were used by the majority of the interventions (13, 93%) to reinforce face-to-face sessions. Projects focused on

educational tasks, such as curriculum design and evaluation (Steinert & McLeod, 2006; Wilkerson *et al.*, 2006), the implementation of an educational innovation (Burdick *et al.*, 2010; Peters, 2002; Robins *et al.*, 2006) or the implementation of a faculty development intervention (Rosenbaum *et al.*, 2005). Small group discussions were used in 11 interventions (79%) to review projects, discuss topics of common interest and share professional stories. Experiential learning methods were used by ten interventions (71%) to provide structured practice opportunities, including storytelling (Robins *et al.*, 2006), participation in the design and delivery of faculty development sessions (Steinert & McLeod, 2006), and the planning and delivery of a session for fellow program participants (Gruppen *et al.*, 2003). Less than half of the interventions (6, 43%) were either developed in response to a stated need (Peters *et al.*, 2002; Sullivan *et al.*, 2006) or explicitly used needs assessments in the design of their curricula and evaluation measures.

Methodological Quality of the Studies

In this section, we shift our focus from describing the 14 interventions in this category to considering the methodological quality of the 14 studies.

Study Goal and Theoretical Framework

Only six of the 14 studies (43%) stated their objective. Some studies described more specific objectives, outlining a particular study question such as identifying the program participants' faculty development activities (Rosenbaum *et al.*, 2005) or exploring participants' perceptions of the utility and personal impact of the intervention (Burdick *et al.*, 2010). In addition, we classified four studies (29%) as citing the relevant literature comprehensively, nine (64%) as doing so adequately, and one (7%) as being quite limited. Four studies (29%) described a conceptual or theoretical framework and three (21%) provided an explicit definition or operationalization of leadership.

Study Design

Six studies (43%) were primarily quantitative (Gruppen *et al.*, 2003; Muller & Irby, 2006; Peters *et al.*, 2002; Sullivan *et al.*, 2006, 2009; Wilkerson *et al.*, 2006), two (14%) were primarily qualitative (Robins *et al.*, 2006; Searle *et al.*, 2006), and six (43%) employed a mixed methods

design. All 12 studies with a quantitative component employed a quasi-experimental design using a single group with no comparison. Seven used a pre-test/post-test design (three retrospective) and five used post-test measures only. Of these 12 studies, three used an immediate post-test only, two used a delayed post-test only, five used both immediate and delayed post-tests, and two did not specify timing of the post-test. Of the eight studies that included a qualitative component, none specified a particular study design.

Data Collection Methods, Sources and Analyses

Author-designed questionnaires were again the most popular method of data collection, used by 12 studies (86%). Sullivan *et al.* (2006) used expert review to validate their questionnaire and Cronbach's alpha was reported as a measure of internal consistency in two studies (Burdick *et al.*, 2010; Sullivan *et al.*, 2006); however, most questionnaires did not appear to have undergone a formal validation process. In addition, most questionnaires were analyzed quantitatively, though Searle *et al.* (2006) reported that they used a constant comparative analysis to generate categories and themes and Simpson *et al.* (2004; 2006) reported using content analysis to analyze a portion of their questionnaire. Four studies (29%) used documentary sources, including CVs, and two (14%) used interviews. All data were collected from program participants.

Study Quality and Strength of Findings

As stated in the previous section, we rated each study on a five-point scale (1 = low; 5 = high). Across raters, the mean rating for study quality in this section was 2.9 and the mean rating for strength of findings was 2.9.

<u>Reported Outcomes</u>

While a number of these interventions reported outcomes in other categories (e.g. teaching, research), we report here on leadership-related outcomes only. Nine educational development interventions (64%) assessed **_reaction_**, which included participant satisfaction, perception of program usefulness and acceptability and value of the activity. Nine interventions (64%) assessed **_learning_**, which included both changes in attitudes (5, 36%) and changes in leadership-related knowledge or skills (6, 43%). Thirteen interventions (93%) had **_behavioural_** outcomes, with nine interventions (64%) identifying *self-reported* behaviour changes and 8 interventions

(57%) identifying *observed* behaviour changes. Four interventions (29%) reported outcomes at the ***results*** level in terms of changes in organizational practice.

Level 1 – REACTION: As in the previous category, satisfaction was usually measured on a 4-5 point Likert scale. Participants rated the programs highly (Muller & Irby, 2006; Sullivan *et al.*, 2006) and valued opportunities for networking (Rosenbaum *et al.*, 2006; Searle *et al.*, 2006; Simpson *et al.*, 2004) as well as the collegial, supportive quality of interactions with their colleagues (Burdick *et al.*, 2010; Srinivasan *et al.*, 2007). Participants in two different interventions described the experience as "transformative" (Steinert & McLeod, 2006; Sullivan *et al.*, 2005).

Level 2a – LEARNING: Outcomes at this level assessed attitudes, which were measured using both self-report questionnaires and interviews. Participants reported an increase in their confidence to assume educational leadership roles (Sullivan *et al.*, 2005; Wilkerson *et al.*, 2006) and in their perceptions of the importance of the different curricular topics covered in the intervention (Burdick *et al.*, 2010; Norcini *et al.*, 2005; Steinert *et al.*, 2003b). Additional outcomes included a broader awareness of learners' needs, themselves as teachers, and medical education (Srinivasan *et al.*, 2007), increased confidence and commitment to medical education, and a sense of belonging to a community (Steinert & McLeod, 2006).

Level 2b – LEARNING: Most learning-related outcomes at this level included the development of knowledge related to teaching. Participants in some interventions also reported an increase in knowledge and skills related to leadership, specifically in the areas of organizational change management and quality improvement, conflict resolution and negotiation, and educational leadership and change (Burdick *et al.*, 2010; Muller & Irby, 2006; Peters *et al.*, 2002; Simpson *et al.*, 2006; Srinivasan *et al.*, 2007; Wilkerson *et al.*, 2006; Wilson & Greenberg, 2004).

Level 3a – BEHAVIOUR (self-reported): Self-reported behavioural changes were identified through both questionnaires and interviews. Nearly all of the participants in Burdick *et al.*'s (2010) study reported applying new leadership-related concepts and skills to the workplace, and over half reported a change in personal leadership style as a result of the program. Participation

in the program also led to new opportunities for leadership and collaboration (Srinivasan, 2007) and many participants achieved (or partly achieved) self-identified goals in areas of administration, leadership, organizational change and building coalitions (Simpson *et al.*, 2004; Sullivan *et al.*, 2006). In addition, participants designed faculty development activities for their peers, developed courses and curricula, implemented educational innovations, became active in educational committees, and reported being viewed as educational leaders in their own departments (Peters *et al.*, 2002; Steinert *et al.*, 2003b; Searle *et al.*, 2006). Interestingly, Robins *et al.* (2006) stated that participants reported that "the program conferred on them legitimacy for their roles as educators, which enabled them to implement innovations more easily" (p. 983). Participants also reported increases in peer-reviewed publications and presentations, grants, pursuit of advanced studies and academic promotions (Norcini *et al.*, 2005; Steinert *et al.*, 2003b; Wilson & Greenberg, 2004).

Level 3b – BEHAVIOUR (observed): Observed behaviour changes were mainly assessed through CV analyses. From pre to post-program, there was an increase in administrative and educational leadership positions (Rosenbaum *et al.*, 2005; Simpson *et al.*, 2004, 2006) as well as academic promotions (Gruppen *et al.*, 2003; Simpson *et al.*, 2004; Sullivan *et al.*, 2009). Participants assumed new leadership roles in medical education (Wilkerson *et al.*, 2006) and new educational responsibilities, including curriculum planning and implementation, developing/chairing courses for medical students and residents, participating in or leading educational committees at the local, regional and national levels, developing new teaching materials, and designing faculty development activities for their colleagues and others at regional and national levels (Gruppen *et al.*, 2003; Robins *et al.*, 2006; Rosenbaum *et al.*, 2005; Simpson *et al.*, 2004; Steinert & McLeod, 2006; Wilkerson *et al.*, 2006). Many programs reported an increase in grants and peer-reviewed presentations and publications (Gruppen *et al.*, 2003; Rosenbaum *et al.*, 2005; Simpson *et al.*, 2006; Steinert & McLeod, 2006; Sullivan *et al.*, 2009) as well as the pursuit of advanced studies (Steinert & McLeod, 2006). Sullivan *et al.* (2009) reported that, as an outcome of participation, program participants assumed leadership roles in all major medical organizations related to palliative care and national end-of-life initiatives.

Level 4a – RESULTS: Few studies examined change at the organizational/systems level. Gruppen *et al.* (2003) reported direct and indirect stimulation of innovations in the medical school, including the introduction of a medical education elective and a medical education grants program. Rosenbaum *et al.* (2006) stated that the departments that had several participants in their program developed more formalized systems and committees to address educational needs and issues within their departments. Participants in this intervention also reported that education now has a higher profile in the department. Additional system changes included new faculty development programs and committee structures (Simpson *et al.*, 2004) and the development of a postgraduate fellowship program and master's program focused on health professions education (Steinert & McLeod, 2006).

CATEGORY 3: LEADERSHIP AS A COMPONENT OF AN ACADEMIC CAREER DEVELOPMENT PROGRAM

Description of the Articles

Eight articles focused on the development of leadership within the context of enhancing the academic success of faculty members. The publication dates of these articles ranged from 1997 to 2008, with the majority of articles (6, 75%) published since 2002. One intervention, the University of California San Diego National Center of Leadership in Academic Medicine (NCLAM) was reported in two studies (Garman *et al.*, 2001; Wingard *et al.*, 2004). Garman *et al.* (2001) reported results from the first two cohorts of the program, while Wingard *et al.* (2004) reported results from the first four cohorts of the program. Though these populations overlap, the study designs differed; thus, these studies were rated separately for methodological quality. In summary, this section includes eight studies that described seven interventions.

Description of the Interventions

Setting

Six of the seven career development interventions (86%) took place in the United States and one (14%) took place in Canada (Talbot *et al.*, 1997). The majority of interventions (6, 86%) were "local" in nature, while one was a national program that welcomed academic and community physicians from across Canada (Talbot *et al.*, 1997).

Program Participants

These interventions primarily targeted junior faculty members, with an emphasis on academic family physicians (Morzinski & Simpson, 2003; Talbot *et al.*, 1997; Woods, 2002). One program (Talbot *et al.*, 1997) welcomed community physicians wanting to move into academic careers. None of these programs included other health science professionals. In two programs (Garman *et al.*, 2001; Pololi *et al.*, 2002), participants volunteered to participate, and one program sought participants by application (Talbot *et al.*, 1997). Selection procedures were not available for the other four interventions, and we did not find any information to suggest that they were compulsory. The number of participants per intervention ranged from 6 to 32, with a mean of 14 participants per cohort. Most evaluation studies reported results from multiple cohorts. The smallest number of participants in an evaluation study was 5 (Howell *et al.*, 2008) and the largest was 67 (Wingard *et al.*, 2004).

Focus of the Intervention

The primary objective of these interventions was to enhance the career development of faculty. Specific objectives included facilitating the career advancement of junior faculty (Garman *et al.*, 2001) and increasing productivity and retention in academics (Morzinski & Simpson, 2003). Leadership objectives included: augmenting institutional and national leadership roles (Howell *et al.*, 2008); nurturing junior faculty to become the next generation of academic leaders (Thorndyke *et al.*, 2006); and developing a team of leaders in family medicine (Talbot *et al.*, 1997). Three interventions (43%) focused additionally on teaching skills, five (71%) on research, two (29%) on communication skills, and three (43%) on using technology to support oral and written communication.

Five interventions addressed leadership explicitly, by including topics such as career planning (3), delivering effective presentations (3), interacting with others (3), self-management (2), negotiation and conflict resolution (2), the nature of change in organizations (2), team building (2), leadership styles (1), value clarification (1), gender and power issues (1), effective meetings (1), time management (1), stress management (1) and quality management (1).

Program Type and Duration

Interventions were classified according to the authors' terminology. Two interventions (29%) were described as one year, part-time fellowships (Talbot *et al.*, 1997; Woods, 2002). The remaining five interventions were described only as programs, lasting from six months to two years. All interventions included multiple components such as face-to-face sessions, mentoring, and individual project work. Sample titles include: National Center for Leadership in Academic Medicine (Garman *et al.*, 2001; Wingard *et al.*, 2004); Scholarly Support Program (Howell *et al.*, 2008); Faculty Development Program (Morzinski & Simpson, 2003); Collaborative Mentoring Program (Pololi *et al.*, 2002); Junior Faculty Development Program (Thorndyke *et al.*, 2006); Faculty Development Fellowship (Woods, 2002); and Five Weekend National Family Medicine Fellowship (Talbot *et al.*, 1997).

Instructional Methods

Most interventions utilized a variety of instructional methods. All interventions included one or more projects that focused on education, clinical practice, and/or research objectives (Morzinski & Simpson, 2003; Thorndyke *et al.*, 2006). One intervention used group projects based on program themes (Talbot *et al.*, 1997), and another intervention (Howell *et al.*, 2008) required participants to both lead their own projects and participate as team members on their colleagues' projects. Four interventions (57%) assigned a senior mentor who provided guidance on career development and projects to each participant, and one intervention (Pololi *et al.*, 2002) was based on a peer-group, collaborative mentoring strategy. Innovative program components included the use of role plays for simulating administrative scenarios (Morzinski & Simpson, 2003) and effective power relationships (Pololi *et al.*, 2002). Four interventions (57%) were either developed in response to a stated need (Howell *et al.*, 2008; Pololi *et al.*, 2002; Thorndyke *et al.*, 2006) or explicitly used needs assessments in the design of their curricula and evaluation measures (Talbot *et al.*, 1997).

Methodological Quality of the Studies

In this section, we shift our focus to specific aspects of the eight studies related to career development.

Study Goal and Theoretical Framework

Seven of the eight studies (88%) stated their objectives. Some studies described more specific objectives, outlining a particular study question such as evaluating whether a formal mentoring process had an impact on junior faculty's self-efficacy (Wingard *et al.*, 2004) or assessing the critical components and global effects of the fellowship (Woods, 2002). We classified two studies (25%) as citing the relevant literature comprehensively and six (75%) as doing so adequately. Three studies (38%) placed their work within a conceptual framework, but none of the studies provided an explicit definition or operationalization of leadership.

Study Design

Five of the eight studies (63%) were primarily quantitative, one (13%) was primarily qualitative, and two (25%) employed a mixed methods design. All seven studies with a quantitative component utilized a quasi-experimental design. Six used a single group design and one (Garman *et al.*, 2001) used a non-equivalent control group. Five used a pre-test/post-test design (one retrospective) and two used post-test measures only. Of the three studies that included a qualitative component, one used a grounded theory design (Woods, 2002); the others did not specify a study design.

Data Collection Methods and Sources

Questionnaires were the most popular method of data collection and most questionnaires were designed by the study authors. Cronbach's alpha was reported as a measure of internal consistency in two studies (Garman *et al.*, 2001; Wingard *et al.*, 2004) and Morzinski and Simpson (2003) used a comprehensive model of faculty development to validate their questionnaire. In total, six studies (75%) used self-report questionnaires, three (38%) used interviews, and three (38%) used documentary sources. Seven studies (88%) collected data from program participants and two studies (25%) collected additional data from colleagues. Howell *et al.* (2008) did not state how their evaluation data were collected.

Study Quality and Strength of Findings

Across raters, the mean rating for study quality in this section was 2.8 and the mean rating for strength of findings was 3.0.

Reported Outcomes

Five career development interventions (71%) assessed **reaction,** which included participant satisfaction, perception of program usefulness and acceptability and value of the activity. Six interventions (86%) assessed **learning,** which included changes in attitudes (4, 57%) as well as changes in knowledge or skills (5, 71%). All interventions assessed **behaviour** changes. Three interventions (43%) reported **self-reported** behaviour changes, while four interventions (57%) reported **observed** behaviour changes. None of the interventions reported outcomes at the **results** level. As in the previous section, we focus here on leadership-related outcomes only.

Level 1 – REACTION: Based on both questionnaires and interviews, program attendance rates were consistently high (Morzinski & Simpson, 2003; Pololi *et al.*, 2002). Participants also rated the programs very positively (Morzinski & Simpson, 2003; Pololi *et al.*, 2002; Talbot *et al.*, 1997; Thorndyke *et al.*, 2006). Thorndyke *et al.* (2006) reported that all participants felt that the program was a valuable educational experience, while both Pololi *et al.* (2002) and Woods (2002) reported that participants valued the camaraderie and sense of community that developed with other participants. In addition, programs were viewed to have an impact on participants' personal effectiveness (Pololi *et al.*, 2002) and career advancement (Thorndyke *et al.*, 2006).

Level 2a – LEARNING: Through questionnaires and interviews, participants reported increased confidence in their administrative skills (Woods, 2002) and a sense of preparedness to fulfill institutional responsibilities (Thorndyke *et al.*, 2006). Compared to their peers who did not participate in the program, program participants in Garman *et al.*'s study (2001) showed statistically significant increases in self-efficacy (e.g. identifying professional goals), personal leadership abilities, and administrative skills. Participants in Pololi *et al.*'s study (2002) viewed the program as "reinforcing, strengthening, or pivotal in their decisions to stay in academic medicine or at our institution" (p. 383) and promoting self-empowerment and personal transformation.

Level 2b – LEARNING: Changes in knowledge were assessed through questionnaires and interviews. In one study, participants increased their knowledge of the expectations and

requirements for promotion (Howell *et al.*, 2008; Thorndyke *et al.*, 2006); another study demonstrated statistically significant knowledge gains in domains of administration and professional academic skills (Morzinski & Simpson, 2003). Participants also reported skill development with respect to communication, negotiation and conflict resolution, value clarification, and planning skills (Pololi *et al.*, 2002; Thorndyke *et al.*, 2006; Woods, 2002).

Level 3a – BEHAVIOUR (self-reported): Participants reported a variety of leadership activities resulting from the intervention. This included taking on new leadership roles (Woods, 2002) and leading projects and programs in their departments (Talbot *et al.*, 1997). Participants also identified their career goals (Thorndyke *et al.*, 2006) and reaffirmed or changed career paths as a result of the fellowship (Woods, 2002). Talbot *et al.* (1997) reported that over 90% of projects initiated during the fellowship were presented in national or regional peer-reviewed settings.

Level 3b – BEHAVIOUR (observed): Howell *et al.* (2008) and Morzinski and Simpson (2003) reported an increase in both internal and national leadership positions from pre-program to post-program. Participation in the program was also seen to contribute to increased numbers of peer-reviewed publications and presentations (Howell *et al.*, 2008; Morzinski & Simpson, 2003; Pololi *et al.*, 2002) as well as academic promotions and the pursuit of advanced studies (Howell *et al.*, 2008). Morzinski and Simpson (2003) reported that 88% of administrative projects initiated during the program were subsequently implemented. Program participation was also associated with increased likelihood of retention at the institution, as compared to other faculty development programs (Morzinski & Simpson, 2003) and national data (Wingard *et al.*, 2004).

SYNTHESIS OF THE FINDINGS

In summary, this review examined 35 interventions that were described in 48 articles. The earliest article, published in 1985, described an intervention that focused primarily on leadership (Category 1), though articles in which leadership was a component of a larger educational development program (Category 2) accounted for the most recent studies. **Table 4** summarizes the study designs and outcome levels of the three categories of interventions.

The vast majority of these interventions took place in the United States, with only three interventions in Canada, one intervention in the United Kingdom and one international intervention. Most programs were local in nature, involving faculty members working with their colleagues at their home institution, though a significant number of interventions that primarily focused on leadership (Category 1) were regional or national programs.

Interventions targeted a range of medical specialties, most commonly family medicine and pediatrics. Category 1 interventions were more often focused on specific populations (junior faculty, senior faculty, women faculty). Interventions that focused primarily on academic career development (Category 3) by and large focused on junior faculty and were directed at medical professionals only, while interventions in the other categories were often interprofessional in nature. The mean number of participants per cohort ranged from a low of 14 for Category 3 interventions to a high of 36 for Category 1 interventions.

Thirty-six percent of Category 1 interventions were short interventions that lasted 3.5 days or less, while interventions in the other categories were all longitudinal programs that lasted at least six months. In terms of instructional methods, experiential learning methods featured prominently in Categories 1 and 2; projects featured prominently in Categories 2 and 3; and mentoring featured prominently in Category 3. Category 1 interventions most often incorporated needs assessments in the design of their programs.

There were notable differences in the way in which each category of interventions implemented a leadership focus. Category 1 interventions focused on a range of different leadership topics, with a particular focus on conflict management, finance, and other interpersonal aspects of leadership such as people management and team-building. In contrast, interventions in Categories 2 and 3 did not always address leadership explicitly in their curricula. The most common curriculum topic in Category 2 focused on a systems aspect of leadership, organizational change management, while the most popular topics in Category 3 tended to focus on personal aspects of leadership such as career planning.

Turning to aspects of study design and methodology, the majority of studies were quantitative in nature, except in Category 2, where the majority of studies employed mixed-methods designs. Quantitative designs were overall quite weak, with the vast majority involving single-group studies. Qualitative designs were typically not described. In all categories (and designs), there was an overwhelming reliance on the use of self-report questionnaires, most with no stated measures of validity or reliability. Similarly, validation strategies for qualitative data collected through interviews were typically not reported.

In terms of outcomes, participants in all categories were highly satisfied with the programs and tended to display enhanced attitudes towards their institutions and their roles as leaders. Interventions in all categories tended to result in knowledge and skill development related to leadership, though these were particularly pronounced in Category 1. Category 2 had more education-focused outcomes and Category 3 included outcomes related to the development of professional academic skills. Similarly, all categories displayed behavioural outcomes related to leadership, but Category 1 interventions were the only ones to demonstrate that participants had an impact on their institutions, as observed by a third party. Systems-level changes were related to broad level organizational issues in Category 1 and were more likely to be focused around educational innovations in Category 2. Systems-level changes were not reported for Category 3.

VII. DISCUSSION

This review focused on faculty development interventions designed to promote leadership in medical education. Despite the use of weak study designs, which limited the conclusions that could be drawn about outcomes, the literature consistently suggested positive changes in faculty members' knowledge, attitudes and skills following participation in a faculty development activity designed to enhance leadership capability. Impacts on the organization (i.e. the systems in which the participants work) were not typically examined; however, some reported outcomes suggested the potential for institutional impact. In this section, we summarize the outcomes of the review and present preliminary observations about the "key features" of programs reporting positive outcomes. We also make some general observations about the nature of faculty development programs to promote leadership in medical education and suggest avenues for future development in research and practice.

Summary of Outcomes

Despite the constraints alluded to above, the literature tended to support the following outcomes:

High Satisfaction with Faculty Development Programs

Overall satisfaction with faculty development programs was high. Participants consistently found the programs to be useful and of both personal and professional benefit. In addition, they appreciated the instructional methods used, especially those with an experiential component and practical focus that applied directly to their personal contexts.

A Change in Attitudes towards Organizational Contexts and Leadership Roles

Participants reported a positive change in attitudes towards their own organizations as well as their leadership capabilities. Some participants reported an increased awareness of − and commitment to − their institution's vision and challenges, whereas others reported greater self-awareness of personal strengths and limitations, increased motivation and confidence in their leadership roles, and a renewed appreciation of the benefits of professional development. A greater sense of community and appreciation of the benefits of networking were also identified post-intervention.

Gains in Knowledge and Skills

Participants reported increased knowledge of leadership concepts, principles, and strategies (e.g. leadership styles; strategic planning; organizational development). They also described perceived gains in specific leadership skills (e.g. personal effectiveness; interpersonal communication; conflict resolution; change management) as well as an increased awareness of leadership roles in academic settings.

Changes in Leadership Behaviour

Self-perceived changes in leadership behaviour were consistently reported. These included a change in leadership styles, the application of new skills to the workplace (e.g. departmental re-organization; team building; time management; financial re-structuring), the adoption of new leadership roles and responsibilities, and the creation of new collaborations and networks.

Observed changes in leadership behaviour, although more limited in nature, complemented these findings and primarily suggested new leadership positions, both in the home organization and in other institutions. Only one study captured the perspective of non-participants, namely the deans of the host institutions. New initiatives, designed and implemented during the intervention (e.g. business plans), were also described at this level, as were increases in scholarly activity (e.g. peer-reviewed publications and presentations) and academic promotions.

Limited Changes in Organizational Practice

Changes at the systems level were not frequently examined. However, in the few studies that did look at this, participants reported changes in organizational practices (e.g. a shift to mission-based budgeting; an increased emphasis on educational scholarship in promotion and tenure), implementation of specific educational innovations, and increased intra- and inter-departmental collaboration. Greater attention to the educational mission and to educational scholarship was also noted, as was the creation of new leadership development programs.

Summary of "Key Features"

While some studies attempted to identify elements that contributed to the success of their program (Morahan *et al.*, 1998; Morahan *et al.*, 2010), or described the benefits and limitations of particular program features (e.g. Steinert *et al.*, 2003b), no studies systematically teased apart features of faculty development that make it effective. A lack of comparable outcome measures across studies also made this analysis difficult. However, we can make some preliminary observations about those features that may have been associated with positive outcomes.

The Use of Multiple Instructional Methods to Achieve Objectives

The majority of interventions included a wide range of instructional methods (e.g. small group discussions; interactive exercises; role plays and simulations). It appears that each program tried to match their methods to specific objectives and different learning styles. Principles of adult learning (e.g. Knowles, 1980) were also commonly cited as influencing instructional design (Morahan *et al.*, 2010; Norcini *et al.*, 2005; Osborn & DeWitt, 2004; Peters *et al.*, 2002; Pololi *et al.*, 2002; Simpson *et al.*, 2006; Sullivan *et al.*, 2005). Interestingly, however, although some programs referred to adult learning theory as part of their curricula (Steinert *et al.*, 2003b;

Wilson & Greenberg, 2004), they did not explicitly state that it guided the design of the intervention itself.

The Role of Experiential Learning and Reflective Practice

Adults prefer to "learn by doing" (Knowles, 1980; Kolb, 1975) and generally value the opportunity to apply new knowledge and skills to the workplace, practice new skills, and receive feedback from peers. In this review, experiential learning opportunities were used to enhance participant interaction (Gruppen *et al.*, 2003; Leslie *et al.*, 2005; Pololi *et al.*, 2002; Sullivan *et al.*, 2005; Wilkerson *et al.*, 2006), reinforce knowledge and skills (Leslie *et al.*, 2005; Rosenbaum *et al.*, 2005; Searle *et al.*, 2006; Srinivasan *et al.*, 2007), and facilitate the application of knowledge and skills to personal contexts (Burdick *et al.*, 2010; Gruppen *et al.*, 2003; Korschun *et al.*, 2007; Muller & Irby, 2006; Robins *et al.*, 2006; Searle *et al.*, 2006; Pololi *et al.*, 2002). It has been said that reflection is needed in order to benefit from experiential learning (Boud *et al.,* 1985). Several reports indicated that participants valued the opportunity to reflect on personal goals and objectives (Pololi *et al.*, 2002; Steinert & McLeod, 2006; Woods, 2002) as well as the process of learning (Morahan *et al.*, 2010; Srinivasan *et al.*, 2007; Steinert *et al.*, 2003b).

The Use of Individual and Group Projects

Projects conducted during the interventions enabled participants to apply the knowledge and skills that they learned to real-world problems in their personal contexts (Gruppen *et al.*, 2003; McCurdy *et al.*, 2004; Morahan *et al.*, 2010). The use of projects served as a powerful motivational tool that enhanced personal accountability (McCurdy *et al.*, 2004; Simpson *et al.*, 2006) and created visibility for the participants within the larger organization (Morahan *et al.*, 2010). In addition, support provided by mentors was important to ensure that projects were aligned with institutional needs (Simpson *et al.*, 2006) and completed in a timely manner (Simpson *et al.*, 2006; Thorndyke *et al.*, 2006). Deans and administrators frequently noted the benefit of projects to the institution (Dannels *et al.*, 2009; Korschun *et al.*, 2007; Thorndyke *et al.*, 2006), though Morzinksi and Simpson (2003) suggested that an emphasis on local projects may have contributed to a decrease in national leadership responsibilities undertaken during the course of their program.

The Value of Peers and the Development of "Communities of Practice"

A number of reports (e.g. Muller & Irby, 2006; Osborn & DeWitt, 2004; Rosenbaum *et al.*, 2006; Srinivasan *et al.*, 2007; Steinert & McLeod, 2006) commented on the benefits of peer and collegial relationships. As in our previous review (Steinert *et al*., 2006), participants highlighted the value of using peers as role models, the benefits of exchanging information and ideas, and the importance of collegial networks to promote and maintain change. The ability to interact and build relationships with peers from diverse backgrounds was highly valued (Burdick *et al.*, 2010; Korschun *et al.*, 2007; Pololi *et al.*, 2002) and participants frequently commented on the importance of a non-threatening, supportive learning environment (Pololi *et al.*, 2002; Thorndyke *et al.*, 2006; Woods, 2002). In addition, the establishment of "communities of practice" was seen as particularly helpful in reducing isolation and providing support to specific populations such as junior faculty (Pololi *et al.*, 2002) and women (Morahan *et al.*, 2010). Other studies indicated that newly developed relationships persisted beyond the intervention (Leslie *et al.*, 2005; Pololi *et al.*, 2002; Searle *et al.*, 2006) and provided institutional benefits. For example, Korschun *et al.* (2007) reported that participants demonstrated an increased willingness to participate in interdepartmental research, teaching and clinical practice, while Dannels *et al.* (2009) found that deans who sponsored three or more ELAM fellows reported greater institutional impact than those deans who sponsored fewer, lending support to the theory that a critical mass of leaders is necessary to effect organizational change (Kanter, 1993).

The Role of Mentorship

Mentors were used to both facilitate project completion (Sullivan *et al.*, 2006; Morahan *et al.*, 2010; Wilkerson *et al.*, 2006) and the acquisition of professional academic skills (Simpson *et al.*, 2006). The participation of mentors was seen as critically important to the success of one intervention (Gruppen *et al.*, 2003) and particularly effective for junior faculty (Garman *et al.*, 2001; Morzinski & Simpson, 2003; Thorndyke *et al.*, 2006; Wingard *et al.*, 2004). Several interventions also included sessions on mentorship and/or provided opportunities for participants to practice their mentorship skills (Leslie *et al.*, 2005; McDade *et al.*, 2004; Osborn & DeWitt, 2004; Rosenbaum *et al.*, 2005; Wilson & Greenberg, 2004; Woods, 2002). Korschun *et al.* (2007) reported that the mentoring component of their program was of variable success. This

may have been due, in part, to the challenge of recruiting qualified senior faculty mentors (Howell *et al.*, 2008; Muller & Irby, 2006). Several interventions addressed this issue through innovative approaches such as collaborative peer mentoring (Pololi *et al.*, 2002), self-mentoring (Rosenbaum *et al.*, 2005), co-mentoring in which second-year program fellows mentored new fellows (Norcini *et al.*, 2005), and the inclusion of both junior and senior faculty within the intervention (Srinivasan *et al.*, 2007).

Institutional Support

Institutional support was identified as critical to the success of many programs. This was evidenced in a number of ways: through direct funding and commitment to protect participants' time (Gruppen *et al.*, 2003; Morahan *et al.*, 1998); encouragement of faculty to enroll (Thorndyke *et al.*, 2006); and direct participation of senior administrators in the program (Korschun *et al.*, 2007; McCurdy *et al.*, 2004; Thorndyke *et al.*, 2006). In support of this notion, Rosenbaum *et al.* (2005) noted that a primary obstacle to successful implementation of participants' projects was the lack of adequate support in their departments. In multiple ways, this review further underscored the role that the institutional culture can play in promoting change.

Avenues for Future Development

In addition to the "key features" summarized above, this review highlighted several issues that are worthy of further exploration for program design, implementation and evaluation, as well as research in this area.

Grounding Faculty Development in Theory and Evidence

As suggested in our previous review (Steinert et al., 2006), we maintain that faculty development should be grounded in both theory and empirical evidence. Models and principles of teaching and learning should inform the planning and development of interventions as well as research to measure outcomes and analyze effects (Mann, 2002). It is interesting to note that few reports in this review situated their activities within a theoretical framework. Russon and Reinelt (2004) make a similar observation in their scan of 55 community leadership development programs. They found that programs did not articulate a program theory or "theory of change" to describe

"how and why a set of activities are expected to lead to outcomes and impacts" (p. 105). There is clearly a need to identify and describe the conceptual frameworks that guide our work in this area. More recently, the idea of learning through participation in communities of practice has been actively discussed (Boud & Middleton, 2003; Lave & Wenger, 2002); this notion, too, has important implications for faculty development designed to promote leadership.

Defining the Nature of Leadership

Grounding this work in the broader literature on leadership (e.g. Bolman & Deal, 2008; George, 2003) would also seem important. In conducting this review, we noted a number of striking differences in how the term "leadership" was characterized and implemented across studies, and indeed, whether it was defined at all. The cited reports tended to use the terms management and leadership interchangeably, ignoring the historical dialectic between the two (Kotter, 1990). There also appeared to be further confusion between what may be called *executive* leadership, *educational* leadership, and *academic* leadership. As a first step, it would seem that articulating a program's notion of leadership – and aligning both program objectives and outcomes with these notions – would be beneficial. A thematic analysis of the leadership concepts addressed in these interventions would also be helpful in identifying similarities and differences across programs.

To advance the field, it may well be timely for us to work towards a meaningful consensus on what we mean by "leadership" in medical education. McDade *et al.* (2008) noted that "the heart of any field is a definition of the phenomena under investigation. Yet, neither scholars nor practitioners of leadership use a common definition" (p. 76). Spencer and Jordan (2001) have identified "vision, an understanding of the educational process, and the ability to implement change" (p. ii38) as the benchmarks of effective leadership to promote educational change in medicine, while Morrison and Jackson (2009) note that leadership is a "quality" that can be found at all levels of a health organization. Some authors talk about leadership attributes such as charisma and integrity (Bennis, 1998; Gardner, 2000), whereas others view leadership as an observable set of practices that can be learned (Kouzes & Posner, 2002). Most of the articles in this review focused on skills and competencies; however, the role of personal attributes and characteristics in leadership development should not be ignored. In addition, examining recent

developments in evidence-based leadership (e.g. Avolio *et al*., 2009) may also be instructive in helping us to design, develop and evaluate faculty development interventions in this area.

Understanding the Role of Context

Though the role of context emerged as a key component in our previous review, few of the studies in this review identified context as a critical feature. In fact, most of the reported interventions were developed to meet the needs of a specific group; as a result, the many positive findings may, in part, reflect congruence between the program design and the needs of the specific population. While we did not observe striking differences in outcomes based on whether a program was offered locally or nationally, or whether participants worked with their own colleagues or not, we suggest that further study into the role of context would be beneficial for two reasons: to assist those who might wish to replicate successful interventions and to provide clarification on how and under what conditions an intervention worked (Cook *et al.,* 2008).

The role of context also seems to be an important factor in looking at those interventions that focused primarily on leadership versus those for whom leadership was a component of a more comprehensive program. Based on the findings in this review, it appears that situating leadership in a specific educational context facilitated educational innovation and change. Participants who were selected for an intervention because they wanted to develop a specific teaching or educational innovation were also reported to be "primed, eager, and ready to learn" (Peters *et al,* 2002, p. 1127).

Exploring the Value of Extended Programs and Follow-Up Sessions

In our previous review, we noted that longer programs, extended over time (e.g. seminar series), tended to produce outcomes not apparent in one-time interventions (e.g. short courses or workshops); that is, these interventions, as well as fellowships, reported more involvement in educational activities following the faculty development activity, implying sustainability over time (Steinert *et al*., 2006). Interestingly, in this review, we saw a preponderance of longitudinal programs that interspersed intensive face-to-face sessions with longer-term individual or group projects. This was a particular trend for the most recent programs, suggesting that this may be a response to the earlier literature.

However, despite the preponderance of extended programs, there was a noticeable lack of program follow-up, a component which appears to be critical in the context of leading change. Interestingly, Leslie *et al.* (2005) reported that the most frequent complaint from program participants was the lack of follow-up after the intervention. Muller and Irby (2006) also commented that program follow-up was an ongoing challenge, and participants in Steinert *et al.*'s study (2003a) requested booster sessions on delegation and conflict management skills. The ELAM program was the only intervention to offer a formal follow-up strategy, which consisted of membership in a society of other ELAM graduates (Morahan *et al.*, 2010).

Promoting the Use of "Alternative" Practices

The current literature demonstrates a reliance on traditional face-to-face methods of instruction. Whereas these methods present advantages in terms of ease of scheduling and building a community of interested educators, we should probably consider other methods that include online and self-directed learning. For example, Fox *et al.* (2001) suggest that successful online interventions require three components: a curriculum, a structure for recording learning and a mechanism such as a facilitated discussion forum to guide participants through the learning cycle. A greater emphasis on peer coaching (Flynn *et al*., 1994) and mentorship (Morzinski *et al*., 1996), as key components of leadership development, would also be warranted, as might the idea of working with teams rather than individuals when addressing leadership competencies. Danzig (2009) has written about the use of story and narrative to teach leadership; future faculty development programs might wish to incorporate story-telling as a specific instructional technique to connect theory and practice – and to capture the personal and complex nature of leadership.

Observations Regarding Methodological Issues

As in our previous review, a number of methodological issues were raised, and we propose consideration of the following:

Promoting More Rigorous Study Designs and Methods

Most of the reports in this review used descriptive, single-group designs to examine outcomes. In fact, only two studies compared participants to non-participants (Dannels *et al.*, 2008; Garman *et al.*, 2001). Single-group designs are problematic because they confound the ability to attribute outcomes directly to the intervention. For example, outcomes such as increased publication rates or enhanced leadership responsibilities may be attributable to many other factors, including natural career progression. The prevalence of single-group designs is also concerning as a number of the studies used either post-intervention measures only and/or collected data several years after the intervention took place. The lack of comparison groups has also made it impossible to subject the results of our analysis to established quantitative techniques such as meta-analysis; instead we have had to rely primarily on descriptive analysis. We also perceived an under-utilization of rigorous qualitative methodologies. The majority of authors did not describe the nature of their qualitative design, and in fact, tended to either confound their method of data collection with the notion of design or failed to describe validation strategies employed, such as member-checking, prolonged engagement in the field, and maintaining an audit trail (Creswell, 2003).

In our previous review, we noted the need to evaluate faculty development programs more systematically and to use sound qualitative and quantitative designs and methods to document outcomes. This recommendation is equally important in this context. Our previous review also suggested that we consider the use of randomized controlled trials or comparison groups in order to make more generalizable statements about whether faculty development does, indeed, make a difference. Norman (2010) has noted, however, that randomized trials may not serve educational interventions consistently well due to the complexity and frequent interaction among variables that are important in understanding the intervention's effect. Norman's argument is further strengthened by the nature of leadership development – and the challenge of implementing change.

By recognizing the complex nature of faculty development, and leadership development in particular (Drescher *et al.,* 2004), we should try to develop a framework for selecting appropriate methodologies that can capture the process of change. In assessing outcomes in this area, many intervening, mediating variables (e.g. personal attributes; teacher's status and responsibilities)

interact with uncontrollable, extraneous factors, and the systematic use of qualitative methods, or mixed designs, could help to capture the complexity of what occurs *during*, and *following*, faculty development interventions (Steinert *et al.*, 2006). Russon and Reinelt (2004) have postulated that interviews are needed to "gather the thick, rich data that captures the meaning and significance of the leadership development experience for participants" (p. 106). This observation highlights the need for a case study approach to program evaluation as well as the role of narrative to demonstrate change. We must also work to align evaluation methodologies with program goals and educational methods, so that the intervention, and not the evaluation method, determines the outcome.

Tapping Multiple Data Sources and Outcome Measures

The results of this review suggest an over-reliance on self-assessment methods and survey questionnaires to assess change. Moving forward, we should consider the use of alternative data sources to assess outcomes that may be more difficult to detect and measure. For example, the study by Dannels *et al.* (2008) is the only one to systematically assess the views of non-participants, namely deans of the host institutions. Given how leadership plays out in a particular context/environment, the perspectives of multiple stakeholders would be invaluable. Moreover, while self-assessment is an important educational activity for participants, novel assessment methods may be necessary to confirm and demonstrate change.

It should also be noted that, similar to our previous review, the majority of studies in this review used questionnaires for which psychometric properties were not reported. We reiterate the importance of using questionnaires that have already been tested for validity and reliability or working to establish such measures. In addition, and as highlighted in our previous review, we should also try to understand the correlations among different measures of performance so that we may be more efficient in our selection of outcome measures. This would make it possible to directly compare the outcomes of the different interventions, and help us to understand the influence of different pedagogical methods or participant selection processes (e.g. nomination of participants vs self-selection).

The fact that few studies assessed change at the system level is congruent with other scans of the leadership development literature. As Russon and Reinelt (2004) noted, most programs focus on individual outcomes such as knowledge, skills, attitudes and perceptions; changes in behaviour, values and beliefs; and leadership paths. Interestingly, this focus may be in keeping with medicine's historical focus on the individual (Bleakley, 2006). There is, however, a clear need to assess outcomes and impact at the organizational, community, and system level.

Assessing Change over Time

Although a few studies assessed change over time (Leslie *et al.*, 2005; McCurdy *et al.*, 2004; Steinert *et al.*, 2003b), the majority did not compare the same outcomes in their immediate and delayed post-intervention assessments. Looking forward, it would be important to explore the durability of change over time, to examine those factors which help to sustain change, and to assess the value of "booster" sessions or other follow-up pursuits.

Russon and Reinelt (2004) make an interesting distinction between *outcomes* (i.e. changes in attitudes, behavior, knowledge, and skills) and *impact* (i.e. the long-term future social change that a program works to create). Given the role of leadership in creating social change, assessment over time is critical. In addition, many of the outcomes anticipated in a leadership development program take time to emerge. This serves as a further reason for longitudinal studies and the assessment of change over time.

Strengths and Limitations of the Review

In reviewing the literature on faculty development for leadership, we faced a number of challenges. The following strengths and limitations were noted in this review:

The Nature of the Review Question

As leadership is a broad and often ambiguous concept, our first goal was to conceptualize our search as broadly as possible. Our inclusion of three very different categories of leadership programs, which address the different types of leadership roles that faculty members may pursue (e.g. executive; educational; academic) is a particular strength of this review. Our exclusive focus on faculty members in medicine has, however, limited our ability to learn from the

perspectives and experiences of leadership interventions designed for other health care professionals, such as dentists (Comer *et al.*, 2002) and nurses (Smith, 2007).

The Review Process

As stated above, we adopted a comprehensive search strategy, including database searching, hand searching, and soliciting recommendations from experts in the field. The cooperation and willingness of our colleagues to supply manuscripts in progress is a particular strength of this review. The results of our search are limited, however, in that they mainly reflect the North American literature. As noted in previous reviews (Freeth *et al.*, 2002; Koppel *et al.*, 2001), this may reflect a publication bias that prevents a fuller picture of faculty development from an international perspective.

In contrast to our previous review, inter-rater reliability was enhanced by a small review team. However, the lack of an international topic review group might be perceived as a weakness, even though the review team sought to maintain critical reflexivity as individuals and as a research team (Freeth *et al*., 2002). Moreover, although we aimed to be as vigilant as possible about data coding and quality control, we apologize in advance for any errors due to personal biases and misinterpretations of data and hope that these will be brought to our attention, to be corrected in the web edition of this review.

In response to earlier criticisms of the BEME coding sheet's emphasis on methodological issues (Dolmans, 2003), we made specific modifications to document conceptual frameworks related to leadership. Further modifications, to more systematically capture cited theories of learning and evaluation, may also be warranted.

The Nature of the Articles Reviewed

In addition to limitations in study design, the level of detail reported about the intervention and its associated outcome(s) varied greatly between articles. It was often difficult to understand the context of the intervention from the background information provided and to report the methodological aspects of the study with precision. An inconsistent use of terminology sometimes led to different interpretations of the same information. While we acknowledge that

some of our concerns (e.g. lack of description of qualitative data analysis methods employed) may be related to the very real space limitations faced by the authors of these reports, we strongly support the need for greater clarity and precision. It is also worth noting that negative results were rarely reported. This may be due to a possible publication bias towards positive results, which is often a challenge for those engaged in a systematic review.

VII. CONCLUSION

The strength of the current evidence in this area is limited. However, a number of valuable lessons can be learned. Below, we identify specific areas for improvement in order to pave the way for enhanced practice and rigorous, high quality research. Interestingly, a number of these recommendations parallel our previous review (Steinert *et al.,* 2006).

Implications for Practice

Based on the review findings, it appears that we should:

- Define our focus. Leadership is clearly a heterogeneous concept. We need to identify what it includes and what knowledge, skills and competencies our interventions are meant to address.
- Make more deliberate use of theory (particularly theories of leadership and learning) in the design and development of our faculty development programs.
- Build on our strengths. The literature describes a number of successful programs, with common features that appear to be associated with positive outcomes. We should incorporate these elements into program design and implementation. For example, we should continue to offer multiple opportunities for experiential learning and reflection and enhance relevance and application through individual and group projects. We should also consider the role of faculty mentors, peer support and communities of practice to promote learning.
- Consider the importance of context. The organizational culture, the program curriculum, the course faculty and the participants all contribute to a context which is likely to be critical to educational change. We need to understand this context and acknowledge its role in leadership development.

- Continue to develop programs that extend over time, to allow for cumulative learning, practice and growth. We should also incorporate follow-up sessions to promote application and reinforcement of new concepts and ideas as they emerge post-intervention.

- Incorporate notions of work-based learning and communities of practice into the design of our interventions. As we have noted, participants consistently value the opportunity to develop new relationships and networks. This educational outcome should be explicitly considered and supported in any initiative designed to promote leadership.

Implications for Research

In line with our findings, and to advance the field, we should:

- Embed our research studies in appropriate conceptual frameworks, utilizing theory in the interpretation of our results.

- Conduct more rigorous research studies using appropriate methodologies. This will require careful definitions of outcomes, planning for evaluation at the inception of any program, and close collaboration with research colleagues. Qualitative methodologies must also be considered as we try to find a way to corroborate anecdotal observations and capture faculty members' stories, many of which demonstrate personal learning and program outcomes.

- Use multiple research methods and data sources to allow for triangulation of data and valid assessment of outcomes. Existing instruments should be used where appropriate in order to promote comparisons across studies and to improve the quality of research in this area. Where new instruments are needed, it is important to assess and report their validity and reliability.

- Carry out process-oriented studies in order to better understand how change occurs, both as a result of the intervention and within the individual (e.g. how did faculty members' attitudes and values change; did the intervention result in enhanced reflective skills). At the same time, we should expand the focus of outcome-oriented studies to compare how different faculty development interventions promote change in faculty competence and performance.

- Develop means of assessing the impact of faculty development on the organization/institution in a more rigorous and systematic fashion.
- Assess change over time in order to understand which interventions or factors may be associated with sustained change. Longitudinal follow-up may also help us to understand how leadership develops throughout a faculty member's career trajectory.
- Collaborate with colleagues within and outside of medicine to promote a more rigorous research agenda. For example, there is much for us to learn from colleagues in the field of management and education. Local research teams and regional networks can also be instrumental in developing – and implementing – a collaborative research agenda that is grounded in practice.

The aim of the Best Evidence Medical Education Collaboration is to encourage teachers to think more clearly about their actions and to utilize relevant evidence to inform their decisions (Harden *et al.*, 1999). The goal of this review was to assess the effect of faculty development activities on faculty members' leadership capabilities and on the organizations in which they work. The literature describes a number of innovative faculty development programs in this area. Participants value these activities and report changes in attitudes, skills, and behaviour. Moreover, despite methodological limitations, certain program characteristics appear to be associated with positive outcomes. Further research is required to explore these associations and to document outcomes at both the individual and organizational level.

Educational and health care change requires leadership (Spencer & Jordan, 2001). We must continue to develop and systematically evaluate professional development activities in this area as we prepare our faculty members to cope with complexity and change. We must also work towards creating a culture of leadership while remembering that leadership, management and organizational development are *not* an end in themselves. As LeMay and Ellis (2008) have said, they are the means to improving the design and delivery of medical education and practice.

ACKNOWLEDGEMENTS

REFERENCES

Articles included in the review findings are marked with an asterisk (*).

*ALUISE, J.J., BOGDEWIC, S.P. & MCLAUGHLIN, C.P. (1985) Organizational development in academic medicine: An educational approach. *Health Care Management Review*, 10(1), pp. 37-43.

AVOLIO, B., WALUMBWA, F. & WEBER, T. (2009) Leadership: Current theories, research, and future directions. *Annual Review of Psychology*, 60, pp. 421-449.

*BACHRACH, D.J. (1997) Developing physician leaders in academic medical centers. *Medical Group Management Journal*, 44(1), pp. 34-38.

BENNIS, W. (1998) *On Becoming a Leader* (London, Arrow).

BLEAKLEY, A. (2006) Broadening conceptions of learning in medical education: The message from teamworking. *Medical Education*, 40(2), pp. 150-157.

BOGDEWIC, S.P., BAXLEY, E.G. & JAMISON, P.K. (1997) Leadership and organizational skills in academic medicine. *Family Medicine*, 29(4), pp. 262-265.

BOLMAN, L.G. & DEAL, T. (2008) *Reframing Organizations: Artistry, Choice, and Leadership, 4th edition* (San Francisco, CA, Jossey-Bass).

BOUD, D., KEOGH, R. & WALKER, D. (1985) What is reflection in learning? in: D. Boud, R. Keogh & D. Walker (Eds) *Reflection: Turning Experience into Learning,* pp. 7-17 (London, Kogan Page).

BOUD, D.& MIDDLETON, H. (2003) Learning from others at work: Communities of practice and informal learning. *Journal of Workplace Learning,* 15(5), pp.194-202.

*BURDICK, W.P., DISERENS, D., FRIEDMAN, S.R., MORAHAN, P.S., KALISHMAN, S., EKLUND, M.A., MENNIN, S. & NORCINI, J.J. (2010) Measuring the effects of an international health professions faculty development fellowship: The FAIMER Institute. *Medical Teacher*, 32(5), pp. 414-421.

*BURDICK, W.P., MORAHAN, P.S. & NORCINI, J.J. (2006) Slowing the brain drain: FAIMER education programs. *Medical Teacher*, 28(7), pp. 631-634.

CENTRA, J.A. (1978) Types of faculty development programs. *Journal of Higher Education,* 49, pp. 151-162

CENTRE FOR REVIEWS AND DISSEMINATION (2009). *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. University of York, UK. [Available at: http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf.]

COOK, D.A., BORDAGE, G., & SCHMIDT, H.G. (2008) Description, justification and clarification: A framework for classifying the purposes of research in medical education. *Medical Education,* 42(2), pp. 128-133.

*COLEMAN, M.T., HEADRICK, L.A., LANGLEY, A.E. & THOMAS, J.X., JR. (1998) Teaching medical faculty how to apply continuous quality improvement to medical education. *Joint Commission Journal on Quality Improvement*, 24(11), pp. 640-652.

COMER, R. W., HADEN, N. K., TAYLOR, R. L. & THOMAS, D. D. (2002) Leadership strategies for department chairs and program directors: A case study approach. *Journal of Dental Education,* 66(4), pp. 514-519.

CRESWELL, J. (2003) *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (Thousand Oaks, CA, Sage Publications).

*DANNELS, S.A., MCLAUGHLIN, J., GLEASON, K.A., MCDADE, S.A., RICHMAN, R. & MORAHAN, P.S. (2009) Medical school deans' perceptions of organizational climate: Useful indicators for advancement of women faculty and evaluation of a leadership program's impact. *Academic Medicine*, 84(1), pp. 67-79.

*DANNELS, S.A., YAMAGATA, H., MCDADE, S.A., CHUANG, Y.-C., GLEASON, K.A., MCLAUGHLIN, J.M., RICHMAN, R.C. & MORAHAN, P.S. (2008) Evaluating a leadership program: A comparative, longitudinal study to assess the impact of the Executive Leadership in Academic Medicine (ELAM) Program for women. *Academic Medicine*, 83(5), pp. 488-495.

DANZIG, A. (2009) How might leadership be taught? The use of story and narrative to teach leadership. *International Journal of Leadership in Education*, 2(2), pp. 117-131.

DRESCHER, U., WARREN, F. & NORTON, K. (2004) Towards evidence-based practice in medical training: Making evaluations more meaningful. *Medical Education,* 38(12), pp. 1288-1294.

*DUDA, R.B. (2008) The impact of a leadership development course for junior academic medical faculty. *Leadership in Health Services*, 21(3), pp. 200-207.

EGGER, M., DAVEY SMITH, G. & ALTMAN, D. G. (Eds) (2001) *Systematic Reviews in Health Care: Meta-Analysis in Context* (London, BMJ Publishing Group).

FLYNN, S.P., BEDINGHAUS, J. SNYDER, C., & HEKELMAN, F. (1994) Peer coaching in clinical teaching: A case report. *Educational Research and Methods,* 26(9), pp. 569-570.

*FOX, N., O'ROURKE., A., ROBERTS, C. & WALKER, J. (2001) Change management in primary care: Design and evaluation of an internet-delivered course. *Medical Education*, 35, pp. 803-805.

FREETH, D., HAMMICK, M., KOPPEL, I., REEVES, S. & BARR, H. (2002) *A critical review of evaluations of interprofessional education*. HSAP Repository. [Available at: http://repos.hsap.kcl.ac.uk/content/m10123/1.3/]

FREETH, D., HAMMICK, M., KOPPEL, I., REEVES, S. & BARR, H. (2005) *Evaluating interprofessional education: A self-help guide*. HSAP Repository. [Available at: http://repos.hsap.kcl.ac.uk/content/m10140/1.1/]

*FROHNA, A. Z., HAMSTRA, S. J., MULLAN, P. B. & GRUPPEN, L. D. (2006) Teaching medical education principles and methods to faculty using an active learning approach: The University of Michigan Medical Education Scholars Program. *Academic Medicine*, 81(11), pp. 975-978.

GARDNER, J. (2000) *The Nature of Leadership* (San Francisco, CA, Jossey-Bass).

*GARMAN, K.A., WINGARD, D.L. & REZNIK, V. (2001) Development of junior faculty's self-efficacy: Outcomes of a National Center of Leadership in Academic Medicine. *Academic Medicine*, 76(10), pp. S74-S76.

GEORGE, W. (2003) *Authentic Leadership: Rediscovering the Secrets to Creating Lasting Value* (San Francisco, CA, Jossey-Bass).

*GRUPPEN, L. , FROHNA, A.Z., ANDERSON, R.M. & LOWE, K.D. (2003) Faculty development for educational leadership and scholarship. *Academic Medicine,* 78(2), pp. 137-141.

GRUPPEN, L.D., SIMPSON, D., SEARLE, N., ROBINS, L., IRBY, D.M. & MULLAN, P.B. (2006) Educational fellowship programs: Common themes and overarching issues. *Academic Medicine*, 81(11), pp. 990-994.

HARDEN, R.M., GRANT, J., BUCKLEY, G. & HART, I.R. (1999) BEME Guide No. 1: Best Evidence Medical Education. *Medical Teacher,* 21(6), pp. 553-562.

*HOWELL, E., KRAVET, S., KISUULE, F. & WRIGHT, S.M. (2008) An innovative approach to supporting hospitalist physicians towards academic success. *Journal of Hospital Medicine,* 3(4), pp. 314-318.

ISSENBERG, S.B., MCGAGHIE, W.C., PETRUSA, E.R., GORDON, D.L. & SCALESE, R.J. (2005) What are the features and uses of high-fidelity medical simulations that lead to most effective learning? BEME Guide No. 4. *Medical Teacher,* 27(1): 10-28.

KANTER, R. M. (1993) *Men and Women of the Corporation* (New York, NY, Basic Books).

KIRKPATRICK, D.L. (1994) *Evaluating Training Programs: The Four Levels* (San Francisco, CA, Berrett-Koehler Publishers).

KNOWLES, M.S. (1980) *The Modern Practice of Adult Education: From Pedagogy to Andragogy, 2nd ed.* (New York, NY, Cambridge Books).

KOLB, D.A. (1984) *Experiential Learning: Experience as the Source of Learning and Development* (Englewood Cliffs, NJ, Prentice-Hall).

KOPPEL, I., BARR, H., REEVES, S., FREETH, D. & HAMMICK, M. (2001) Establishing a systematic approach to evaluating the effectiveness of interprofessional education. *Issues in Interdisciplinary Care,* 3(1), pp. 41-49.

*KORSCHUN, H.W., REDDING, D., TEAL, G.L. & JOHNS, M.M.E. (2007) Realizing the vision of leadership development in an academic health center: The Woodruff Leadership Academy. *Academic Medicine*, 82(3), pp. 264-271.

KOTTER, J.P. (1990) *A Force for Change: How Leadership Differs from Management* (New York, NY, Free Press).

KOUZES, J.M. & POSNER, B.Z. (2002) *The Leadership Challenge, 3rd Edition* (San Francisco, CA, Jossey-Bass).

LAVE, J. & WENGER, E. (2002) Legitimate peripheral participation in communities of practice, in: R. Harrison, F. Reeve, A. Hanson & J. Clarke (Eds) *Supporting Lifelong Learning, Volume 1* (London, UK, Routledge Farmer).

LEMAY, N. & ELLIS, A. (2008) Leadership can be learned, but how is it measured? *Management Sciences for Health*, 8, pp. 1-29.

*LESLIE, L.K., MIOTTO, M.B., LIU, G.C., ZIEMNIK, S., CABRERA, A.G., CALMA, S., HUANG, C. & SLAW, K. (2005) Training young pediatricians as leaders for the 21st century. *Pediatrics*, 115(3), pp. 765-773.

MANN, K.V. (2002) Thinking about learning: Implications for principle-based professional education. *Journal of Continuing Education in the Health Professions,* 22(2), pp. 69-76.

MAUDSLEY, G. (2011) Mixing it but not mixed-up: Mixed methods research in medical education (a critical narrative review). *Medical Teacher*, 33(2): e92-e104.

*MCCURDY, F.A., BECK, G., MAROON, A., GOMES, H. & LANE, P.H. (2004) The Administrative Colloquium: Developing management and leadership skills for faculty. *Ambulatory Pediatrics*, 4(1 SUPPL), pp. 124-128.

*MCDADE, S.A., RICHMAN, R.C., JACKSON, G.B. & MORAHAN, P.S. (2004) Effects of participation in the Executive Leadership in Academic Medicine (ELAM) program on women faculty's perceived leadership capabilities. *Academic Medicine*, 79(4), pp. 302-309.

*MORAHAN, P.S., GLEASON, K.A., RICHMAN, R.C., DANNELS, S. & MCDADE, S.A. (2010). Advancing women faculty to senior leadership in US academic health centers: Fifteen years of history in the making. *NASPA Journal about Women in Higher Education*, 3(1), pp. 137-162.

*MORAHAN, P.S., KASPERBAUER, D., MCDADE, S.A., ASCHENBRENER, C.A., TRIOLO, P.K., MONTELEONE, P.L., COUNTE, M. & MEYER, M.J. (1998) Training future leaders of academic medicine: Internal programs at three academic health centers. *Academic Medicine*, 73(11), pp. 1159-1168.

MORRISON, S. & JACKSON, N. (2009) Leadership and management in education and training, in: Y. Carter & N. Jackson (Eds) *Medical Education and Training: Theory to Delivery*, pp. 309-320 (London, Oxford University Press).

MORZINSKI, J.A., DIEHR, S., BOWER, D.J. & SIMPSON, D.E. (1996) A descriptive, cross-sectional study of formal mentoring for faculty. *Family Medicine,* 28(6), pp. 434-438.

*MORZINSKI, J.A. & SIMPSON, D.E. (2003) Outcomes of a comprehensive faculty development program for local, full-time faculty. *Family Medicine*, 35(6), pp. 434-439.

*MULLER, J.H. & IRBY, D.M. (2006) Developing educational leaders: The Teaching Scholars Program at the University of California, San Francisco, School of Medicine. *Academic Medicine*, 81(11), pp. 959-964.

*NORCINI, J., BURDICK, W. & MORAHAN, P. (2005) The FAIMER Institute: Creating international networks of medical educators. *Medical Teacher*, 27(3), pp. 214-218.

NORMAN, G. (2010) Is experimental research passé. *Advances in Health Sciences Education,* 15, pp. 297-301.

*OSBORN, L.M. & DEWITT, T. (2004) The HRSA-APA Faculty Development Scholars Program: Executive leadership track. *Ambulatory Pediatrics*, 4(1 SUPPL), pp. 98-102.

*PETERS, A.S., LADDEN, M.D., KOTCH, J.B. & FLETCHER, R.H. (2002) Evaluation of a faculty development program in managing care. *Academic Medicine*, 77(11), pp. 1121-1127.

*POLOLI, L.H., KNIGHT, S.M., DENNIS, K. & FRANKEL, R.M. (2002) Helping medical school faculty realize their dreams: An innovative, collaborative mentoring program. *Academic Medicine*, 77(5), pp. 377-384.

*RICHMAN, R.C., MORAHAN, P.S., COHEN, D.W. & MCDADE, S.A. (2001) Advancing women and closing the leadership gap: The Executive Leadership in Academic Medicine (ELAM) program experience. *Journal of Women's Health and Gender-Based Medicine,* 10(3), pp. 271-277.

*ROBINS, L., AMBROZY, D. & PINSKY, L.E. (2006) Promoting academic excellence through leadership development at the University of Washington: The Teaching Scholars Program. *Academic Medicine*, 81(11), pp. 979-983.

*ROSENBAUM, M.E., LENOCH, S. & FERGUSON, K.J. (2005) Outcomes of a Teaching Scholars Program to promote leadership in faculty development. *Teaching & Learning in Medicine*, 17(3), pp. 247-252.

*Rosenbaum, M.E., Lenoch, S. & Ferguson, K.J. (2006) Increasing departmental and college-wide faculty development opportunities through a Teaching Scholars Program. *Academic Medicine*, 81(11), pp. 965-968.

Russon, C. & Reinelt, C. (2004) The results of an evaluation scan of 55 leadership development programs. *Journal of Leadership & Organizational Studies,* 10(3), pp. 104-107.

Sandelowski, M., Voils, C. I. & Barroso, J. (2006) Defining and designing mixed research synthesis studies. *Research in the Schools, 13*(1), pp. 29-40.

*Searle, N.S., Thompson, B.M. & Perkowski, L.C. (2006) Making it work: The evolution of a medical educational fellowship program. *Academic Medicine*, 81(11), pp. 984-989.

Seavey, J.W. & Hiller, M.D. (1984) Faculty development in health administration: A case study of faculty internships. *Journal of Health Administration Education*, 2(1), pp. 75-81.

Sheets, K.J. & Schwenk, T.L. (1990) Faculty development for family medicine educators: An agenda for future activities. *Teaching and Learning in Medicine*, 2, pp. 141-148.

*Simpson, D.E., Bragg, D., Biernat, K. & Treat, R. (2004) Outcomes results from the evaluation of the APA/HRSA Faculty Scholars Program. *Ambulatory Pediatrics*, 4(1 SUPPL), pp. 103-112.

*Simpson, D., Marcdante, K., Morzinski, J., Meurer, L., McLaughlin, C., Lamb, G., Janik, T. & Currey, L. (2006) Fifteen years of aligning faculty development with primary care clinician-educator roles and academic advancement at the Medical College of Wisconsin. *Academic Medicine*, 81(11), pp. 945-953.

*Sloma-Williams, L., McDade, S.A., Richman, R.C. & Morahan, P.S. (2009) The role of self-efficacy in developing women leaders, in: D.R. Dean, S.J. Bracken & J.K. Allen (Eds) *Women in Academic Leadership: Professional Strategies, Personal Choices,* pp. 50-73 (Sterling, VA, Stylus).

Smith, G.R. (2007) Project LEAD. *The Association of Black Nursing Faculty Journal*, 18(3), pp. 72-73.

Spencer, J. & Jordan, R. (2001) Educational outcomes and leadership to meet the needs of modern health care. *Quality in Health Care*, 10(SUPPL II), pp. ii38-ii45.

*Srinivasan, M., Pratt, D.D., Collins, J., Bowe, C.M., Stevenson, F.T., Pinney, S.J. & Wilkes, M.S. (2007) Developing the master educator: Cross disciplinary teaching scholars program for human and veterinary medical faculty. *Academic Psychiatry*, 31(6), pp. 452-464.

STEINERT, Y., MANN, K., CENTENO, A., DOLMANS, D., SPENCER, J., GELULA, M. & PRIDEAUX, D. (2006) A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Medical Teacher*, 28(6), pp. 497-526.

*STEINERT, Y. & MCLEOD, P.J. (2006) From novice to informed educator: The Teaching Scholars Program for educators in the health sciences. *Academic Medicine*, 81(11), pp. 969-974.

*STEINERT, Y., NASMITH, L. & DAIGLE, N. (2003a) Executive skills for medical faculty: A workshop description and evaluation. *Medical Teacher*, 25(6), pp. 666-668.

*STEINERT, Y., NASMITH, L., MCLEOD, P.J. & CONOCHIE, L. (2003b) A teaching scholars program to develop leaders in medical education. *Academic Medicine*, 78(2), pp. 142-149.

*STOLLER, J.K., BERKOWITZ, E. & BALLIN, P.L. (2007) Physician management and leadership education at the Cleveland Clinic Foundation: Program impact and experience over 14 years. *Journal of Medical Practice Management*, 22(4), pp. 237-242.

*SULLIVAN, A.M., GADMER, N.M. & BLOCK, S.D. (2009) The Project on Death in America Faculty Scholars Program: A report on scholars' progress. *Journal of Palliative Medicine*, 12(2), pp. 155-159.

*SULLIVAN, A.M., LAKOMA, M.D., BILLINGS, J.A., PETERS, A.S., BLOCK, S.D. & PCEP CORE FACULTY (2005) Teaching and learning end-of-life care: Evaluation of a faculty development program in palliative care. *Academic Medicine,* 80(7), pp. 657-668.

*SULLIVAN, A.M., LAKOMA, M.D., BILLINGS, J.A., PETERS, A.S., BLOCK, S.D. & PCEP CORE FACULTY (2006) Creating enduring change: Demonstrating the long-term impact of a faculty development program in palliative care. *Journal of General Internal Medicine*, 21(9), pp. 907-914.

*TALBOT, Y., BATTY, H. & ROSSER, W.W. (1997) Five weekend national family medicine fellowship. *Canadian Family Physician*, 43, pp. 2151-2157.

*THORNDYKE, L.E., GUSIC, M.E., GEORGE, J.H., QUILLEN, D.A. & MILNER, R.J. (2006) Empowering junior faculty: Penn State's faculty development and mentoring program. *Academic Medicine*, 81(7), pp. 668-673.

*WILKERSON, L., UIJTDEHAAGE, S. & RELAN, A. (2006) Increasing the pool of educational leaders for UCLA. *Academic Medicine*, 81(11), pp. 954-958.

*WILSON, M. & GREENBERG, L. (2004) Overview of the Educational Scholarship Track. *Ambulatory Pediatrics,* 4 (1 SUPPL), pp. 88-91.

*WINGARD, D.L., GARMAN, K.A. & REZNIK, V. (2004) Facilitating faculty success: Outcomes and cost benefit of the UCSD National Center of Leadership in Academic Medicine. *Academic Medicine*, 79(10), pp. S9-S11.

*WOODS, S.E. (2002) A qualitative assessment of one cohort from the University of North Carolina Family Medicine Faculty Development Fellowship. *Family Medicine*, 34(2), pp. 126-131.

*WOODS, S.E. & GRIGGS, G.K. (1994) A curriculum for teaching faculty budgeting and financial management skills. *Family Medicine*, 26(10), pp. 645-647.

ZAHER, C.A. (1996) Learning to be a leader. *The Physician Executive*, 22(9), pp. 10-17.