

Evaluation of and Feedback for Academic Medicine Leaders: Developing and Implementing the Memorial Method

James Rourke, MD, CCFP-EM, MCLinSc, FCFP, FRRMS, FCAHS, LLD,
Stephen Bornstein, MA, PhD, Cathy Vardy, MD, FRCPC, David Speed, MASP, PhD,
Tyrone White, and Paula Corbett, BVocEd

Abstract

Problem

Giving and receiving honest and helpful feedback for leadership development is a common challenge in all types of organizations but particularly in academic medicine.

Approach

At Memorial University of Newfoundland, in 2014, a consensus emerged to develop a new method for evaluating the leadership performance of the discipline chairs, dean, and vice dean, and to provide these leaders with the evaluation results to help them improve their performance. The leaders responsible for developing and implementing this

method (called the Memorial Method) decided to use a survey to obtain faculty members' perceptions about their leader's performance. Beginning in October 2014, a portion of several regular meetings of the discipline chairs with the dean and vice dean was used to develop the survey, by first discussing the broad dimensions of leadership performance, then discussing these dimensions in more detail and drafting specific questions. The resulting survey included 44 quantitative questions addressing eight leadership dimensions. In March–April 2015, the survey was administered electronically to full-time faculty members on a confidential basis.

The results were compiled and reported to each discipline chair and to the dean and vice dean.

Outcomes

In total, 144/249 faculty responded to the survey (response rate: 58%). For the various dimensions, individual chairs' mean scores ranged from 2.82 to 4.70, and overall mean scores ranged from 3.57 to 4.24. Psychometric properties of the survey suggested it was both reliable and valid.

Next Steps

The survey will be repeated, this time with part-time as well as full-time faculty included.

Problem

Giving and receiving honest and helpful feedback for leadership development is a common challenge in all types of organizations but particularly in academic medicine. Leaders within academic medicine organizations such as deans, vice deans, and discipline chairs are at the crossroads of education, research, and discipline care. Faculty asked to evaluate a leader may be hesitant about giving honest feedback when the leader they are evaluating is responsible for their career progression and work assignments, and may also be a

colleague with whom they share patient responsibilities in a group practice or an on-call system.¹

Approach

At Memorial University of Newfoundland (Memorial), St. John's, Newfoundland and Labrador, Canada, in fall 2014, a consensus emerged among the leaders of the Faculty of Medicine to develop a new method for evaluating the leadership performance of the discipline chairs, dean, and vice dean (collectively, the academic leaders) and to provide them with the evaluation results to help them improve their performance. All discipline chairs agreed to work with the dean and vice dean on developing and implementing this evaluation and feedback process.

Analysis of the literature on performance assessment provided the academic leaders with guidance on the best general approach to use. Rather than using complex, time-consuming approaches such as 360 reviews, the

academic leaders reached a decision to survey faculty members who report to each leader to obtain their perceptions on the leader's performance.^{2,3} However, the literature review also revealed that none of the existing performance assessment instruments were sufficiently well aligned with the job characteristics of the academic leaders involved or with the administrative context of Memorial's Faculty of Medicine to provide sufficient face validity and relevance.^{4,5} Accordingly, the academic leaders felt that a new survey and a new approach to its use were needed. We refer to this new survey and its implementation as "the Memorial Method." (It is important to note that, when this work was undertaken and completed, the recent *Academic Medicine* article by Palmer and colleagues⁶ on giving feedback to department chairs, which might have proved useful, had not yet been published.)

Creating the survey

The core challenge in developing the survey was to develop a set of questions

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

Correspondence should be addressed to James Rourke, Discipline of Family Medicine, Faculty of Medicine, Memorial University of Newfoundland, Health Sciences Centre, St. John's, NL, Canada, A1B3V6; telephone: (709) 864-6549; e-mail: Jrouke@mun.ca.

Acad Med. XXXX;XX:00–00.

First published online

doi: 10.1097/ACM.0000000000001722

Copyright © 2017 by the Association of American Medical Colleges

to target the leadership dimensions the academic leaders and their faculty would regard as important. An initial discussion about which leadership dimensions were important was held at a regular meeting of the discipline chairs with the dean and vice dean in October 2014. At this meeting, it was agreed that the discipline chairs would work with the dean and vice dean to develop the leadership dimensions on which the survey would focus and the questions that would be used to assess performance on each dimension. To allay confidentiality concerns and encourage faculty participation, the discipline chairs accepted a suggestion by the dean that the entire process be managed by a third party, the Newfoundland and Labrador Centre for Applied Health Research (NLCAHR), a respected research organization at Memorial with an arm's length relationship to the faculty. Throughout the following few months, a portion of several regular meetings of the discipline chairs with the dean and vice dean was set aside to develop the survey. In these meetings, the academic leaders first discussed broad dimensions of leadership performance, then discussed each dimension in more detail and drafted and discussed specific questions to address each. Through this process, a consensus emerged on the most important leadership dimensions to assess and on the questions and methods for doing so.

The initial set of leadership dimensions and draft questions was then revised, refined, and consolidated by the dean in consultation with the director of the NLCAHR. The resulting set of draft questions was then approved by the academic leaders in January 2015; a pilot was run with a select group of faculty members to test the clarity of the questions and to determine the amount of time required to complete the survey. Following the pilot, the dean and the director of NLCAHR revised the questions again to improve clarity. The result was a compact survey containing 44 questions (on a five-point Likert scale, where 1 = strongly disagree, 5 = strongly agree) addressing eight leadership dimensions: communication, scheduling meetings, decision making, strategic leadership, recruitment and retention, support of faculty or staff, engagement, and integrity (List 1). Some

List 1

Survey Used to Evaluate Leadership Performance in Academic Medicine (With 44 Quantitative Questions Addressing Eight Leadership Dimensions),^a the Memorial Method, Memorial University of Newfoundland Faculty of Medicine, March–April 2015

Communication

My discipline chair...

- Is an active listener.
- Is responsive to requests and questions.
- Sends me too many e-mails.
- Sends me too few e-mails.
- Sends me e-mails that are helpful.

Scheduling meetings

My discipline chair...

- Schedules too many meetings.
- Schedules too few meetings.
- Schedules meetings that are helpful.
- Conducts too many one-on-one meetings and/or phone calls with me.
- Conducts too few one-on-one meetings and/or phone calls with me.
- Conducts one-on-one meetings and/or phone calls that are helpful.

Decision making

My discipline chair...

- Usually acts unilaterally for major decisions.
- Usually consults with others before making major decisions.
- Promptly addresses problems.
- Lets problems build up.
- Should delegate tasks more often.
- Delegates tasks too often.
- Micromanages tasks performed by others.
- Gives ambiguous instructions when delegating tasks.

Strategic leadership

My discipline chair...

- Has developed a good collective vision and plan for the discipline.
- Provides transparent leadership.
- Bridges academic and clinical programs.
- Effectively advocates for our program within the Faculty of Medicine, regional health authority, university, government, etc.

Recruitment and retention

My discipline chair...

- Is successful in recruiting new faculty.
- Recruits new faculty in a transparent fashion.
- Supports existing faculty.

Support of faculty or staff

My discipline chair...

- Facilitates goal-specific development of faculty members.
- Provides regular feedback to faculty members on their performance.
- Provides useful feedback to faculty members on their performance.
- Effectively manages staff.
- Is a role model for academic, teaching, and clinical excellence.
- Demonstrates a positive attitude about self, colleagues, and the Faculty of Medicine.
- Provides encouragement, motivation, and support to faculty and staff.
- Recognizes and rewards excellence and gives credit where credit is due.

Engagement

My discipline chair...

- Actively participates in all aspects of the discipline.
- Is available to faculty.
- Is available to staff.
- Is available to students.
- Builds morale and promotes a positive working environment.

Integrity

My discipline chair...

- Demonstrates integrity.
- Is approachable for difficult problems.
- Promotes professionalism in others.
- Applies the rules fairly to everyone.
- Demonstrates compassion or empathy.

^aThe full survey instrument can be found at <http://fluidsurveys.com/surveys/nlcahr/sample/>. All the questions required responses on a five-point Likert scale, where 1 = strongly disagree, 5 = strongly agree. Some questions contained negative statements to counteract possible effects of positivity bias (i.e., the tendency of respondents to simply agree with statements). For analysis, however, responses to negative statement questions were reverse coded so that higher scores always reflected better leadership. Here all questions are listed as beginning with "My discipline chair..." but in surveys that were used to evaluate the dean and vice dean all questions began with "My dean..." and "My vice dean..." respectively.

of the questions contained negative statements to counteract possible effects of positivity bias (i.e., the tendency of respondents to simply agree with statements). For analysis, however, responses to the negative statement questions were reverse-coded so that

higher scores always reflected better leadership.

Administration of the survey

With the survey instrument ready to go, the next set of decisions involved how it was to be implemented. The academic

leaders chose an online, rather than a paper-based, approach to maximize confidentiality, simplify the tracking of responses, and facilitate analysis of the results. After a careful review of available software options, the dean and the director of NLCAHR decided to use the Web-based FluidSurveys platform from SurveyMonkey (Ottawa, Ontario, Canada) because this platform could provide all the capacity required while guaranteeing full confidentiality.

In March 2015, the dean sent an e-mail to all full-time faculty members (except for those in disciplines with fewer than five full-time faculty members, $n = 249$) explaining the purpose of the survey and how it was designed, and alerting them that they would each soon receive an invitation to evaluate their discipline chair. Using FluidSurveys, the NLCAHR then generated an e-mail with the dean's signature describing the key features of the survey and including a link to it. A similar process was used to ask the discipline chairs to fill out surveys on the performance of the dean and vice dean. Participants were given a total of four weeks to complete the survey. FluidSurveys assigned each respondent a numerical identifier to mask their name and e-mail address. As the deadline for responding approached, FluidSurveys automatically sent two e-mail reminders to all potential participants who had not yet completed the survey.

Special considerations

Given the relatively small numbers of respondents and academic leaders involved and the delicacy of the contents of the survey, it was essential to design a process that would provide the highest possible level of anonymity and privacy for both groups. Accordingly, the Memorial Method was designed with the following safeguards:

- the use of a highly regarded, secure, Web platform;
- careful and sustained anonymization of respondents' identities;
- limiting the survey to quantitative responses rather than including places for qualitative responses that might compromise the respondent's identity;
- third-party administration and analysis; and

- the exclusion of disciplines with fewer than five full-time faculty members.

In addition to this attention to confidentiality, the Memorial Method was designed to maximize two other variables—the response rate and the honesty of the responses. Securing respectable response rates was challenging, given the respondents' demanding schedules. The design sought to maximize response rates by emphasizing the confidentiality safeguards, keeping the survey as compact as possible and assuring potential respondents that no more than 10 minutes of their time was needed, and employing survey methodology best practices for communicating with potential respondents (an initial e-mail invitation followed by two reminder e-mails).⁷

To maximize the honesty of the responses, communications with potential respondents emphasized the arm's length role and neutrality of the NLCAHR, the full cooperation of the academic leaders who were being evaluated, and the safeguards that had been put in place to ensure confidentiality.⁸

Reporting the survey results

The NLCAHR conducted all data analyses using the randomly generated numerical identifiers. NLCAHR staff generated reports for each discipline chair, the dean, and the vice dean that reported their performance as compared with their peers overall, by leadership dimension, and by each question. NLCAHR staff also generated a summary report for the dean that reported the performance of each discipline chair and the vice dean both by leadership dimension and overall, as well as his or her *relative* performance, as compared with the mean scores for all discipline chairs. To compare performance among the discipline chairs, NLCAHR staff used Mann–Whitney–Wilcoxon tests with critical dimension reduced ($P = .0057$) to ensure that the family-wise error rate did not exceed $P = .05$.

Outcomes

In total, 144 of the 249 eligible full-time faculty responded to the survey (response

rate: 58%). Response rates for individual disciplines ranged from 43% to 70%. The specific responses submitted contained sufficient variability in the mean scores for the various leadership dimensions to indicate that respondents had taken the time to answer questions thoughtfully rather than simply ticking the same box on all questions. As can be seen in Table 1, individual discipline chairs' mean scores for the various dimensions ranged from 2.82 to 4.70, and the overall mean scores ranged from 3.57 to 4.24. In addition, the lowest scores on many of the questions were low enough to suggest that respondents had not hesitated to provide negative feedback, and thus that social desirability bias (i.e., the tendency to answer questions in a manner that will be viewed favorably by others) had not played a role. To ensure that these variations in scores signified meaningful performance differences, NLCAHR staff performed appropriate statistical tests. Discipline chairs who were statistical outliers were identified, with particular attention being paid to those with comparatively low scores overall or on specific dimensions.

The analysis paid attention to the reliability and validity of the survey. With regard to validity, the eight leadership dimensions and the 44 quantitative questions were sufficiently aligned with the existing literature on leadership and sufficiently similar to existing leadership assessment instruments to suggest both convergent validity and face validity.^{9,10} Moreover, the central role that the discipline chairs played in the development of the survey provides a strong argument for ecological validity. In terms of reliability, each subscale was found to be acceptable (see Table 1), with Cronbach α ranging from 0.66 (scheduling meetings) to 0.92 (integrity). These psychometric properties suggest that the Memorial Method is both reliable and valid.

The design and administration of the survey generated a good response rate and a robust set of credible responses. This allowed the dean to use the survey results in a confident and meaningful way as part of one-on-one discussions with the discipline chairs and vice dean. For each academic leader, including the dean and vice dean, the report highlighted dimensions

in which his or her performance was commendable and dimensions in which improvement was needed. Thus, with the exception of the chairs of the smallest disciplines, the reports provided the dean with quantitative data to add to a variety of other, more traditional, inputs as part of a cumulative multimethod assessment of leadership performance.

Next Steps

To the best of our knowledge, there have been no previous attempts to design a leadership survey in academic medicine using a pool of items that were generated through consultations with the leaders themselves.

One sign that the Memorial Method was well designed was that shortly after the first round of surveys was completed, the associate dean of postgraduate medical education in the Memorial Faculty of Medicine asked the NLCAHR to help administer a similar survey to evaluate the performance of residency program

directors. That survey has recently been undertaken, using slightly revised questions, an enhanced approach to confidentiality, and improved electronic data collection methods. One key modification was that each program director was evaluated by two different groups—the program’s faculty members and its residents. The relative ease with which the initial instrument was revised for a slightly different use suggests that the Memorial Method could be readily adapted for application in multiple environments, both in medicine and in other faculties, and, possibly, in similarly structured nonacademic organizations.

The next step will be to repeat the survey for the academic leaders, this time including part-time as well as full-time faculty members. One other modification being considered for this next implementation is the addition of a self-administered component. The idea would be to have each of the academic leaders being evaluated use the survey to evaluate him- or herself as well. The results

would then be compared with those submitted by faculty members, so that the comparison could provide each academic leader and the dean with useful information for discussion.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: As the survey was deemed to be part of the Faculty of Medicine’s administrative review process and only anonymized and aggregate data were to be included in any publications, ethics approval was not required.

J. Rourke is professor of family medicine and was dean of medicine 2004–2016, Memorial University of Newfoundland, St. John’s, Newfoundland and Labrador, Canada; ORCID: <http://orcid.org/0000-0001-8019-0294>.

S. Bornstein is professor of political science and medicine and director, Newfoundland and Labrador Centre for Applied Health Research, Memorial University of Newfoundland, St. John’s, Newfoundland and Labrador, Canada.

C. Vardy is vice dean of medicine and professor of pediatrics, Memorial University of Newfoundland, St. John’s, Newfoundland and Labrador, Canada.

D. Speed is a research officer, Newfoundland and Labrador Centre for Applied Health Research, and term lecturer in psychology, Memorial University of Newfoundland, St. John’s, Newfoundland and Labrador, Canada.

T. White is manager of finance, awards, and administration, Newfoundland and Labrador Centre for Applied Health Research, Memorial University of Newfoundland, St. John’s, Newfoundland and Labrador, Canada.

P. Corbett is manager of academic affairs, Faculty of Medicine, Memorial University of Newfoundland, St. John’s, Newfoundland and Labrador, Canada.

Table 1

Mean Scores of Discipline Chairs, Using the Memorial Method for Evaluating Leadership Performance in Academic Medicine, Memorial University of Newfoundland Faculty of Medicine, March–April 2015

Discipline chairs	Leadership dimensions ^a								Overall
	1	2	3	4	5	6	7	8	
Coefficient of variation	0.18	0.17	0.19	0.20	0.20	0.19	0.18	0.18	0.16
Cronbach α	0.73	0.66	0.83	0.83	0.76	0.91	0.85	0.92	0.97
Mean score									
Discipline A	3.98	3.78	3.85	4.00	4.28	3.97	4.25	4.53	4.03
Discipline B	4.16	3.90	3.70	3.95	3.93	4.05	4.04	4.20	3.98
Discipline C	3.82	3.78	3.64	3.75	4.00	3.86	3.90	4.10	3.83
Discipline D	4.50	4.38	4.41	4.44	4.50	4.69	4.70	4.70	4.54
Discipline E	4.14	3.82	3.96	3.82	3.97	4.06	4.22	4.46	4.05
Discipline F	3.34	3.57	2.82 ^b	3.72	3.30 ^b	3.64	3.98	3.38 ^b	3.44
Discipline G	3.58	3.47	3.23 ^b	3.93	3.90	3.58	3.75	3.83 ^b	3.61
Discipline H	4.40 ^c	4.12	4.09 ^c	4.36	4.21	4.42 ^c	4.65 ^c	4.60	4.34 ^c
Discipline I	3.28	3.57	3.21	3.35	3.40	3.51	3.44	4.10	3.46
Overall	3.92	3.70	3.57	3.92	3.92	3.92	4.06	4.24	3.86

^aDimensions: 1 indicates communication; 2, scheduling meetings; 3, decision making; 4, strategic leadership; 5, recruitment and retention; 6, support of faculty or staff; 7, engagement; 8, integrity. All dimensions were rated using a five-point Likert scale, where 1 = strongly disagree, 5 = strongly agree. Comparisons are on the vertical axis, with each discipline chair being compared to all other discipline chairs for each leadership dimension. In some smaller disciplines, very high or low values may not be significant because of sample size effect.

^bSignificantly lower than other disciplines, using Mann–Whitney–Wilcoxon tests with critical dimension reduced ($P = .0057$) to ensure that the family-wise error rate did not exceed $P = .05$.

^cSignificantly higher than other disciplines, using Mann–Whitney–Wilcoxon tests with critical dimension reduced ($P = .0057$) to ensure that the family-wise error rate did not exceed $P = .05$.

References

- 1 Craighead PS, Anderson R, Sargent R. Developing leadership within an academic medical department in Canada: A road map for increasing leadership span. *Healthc Q*. 2011;14:80–84.
- 2 Guyatt G, Cook D, King D, Nishikawa J, Brill-Edwards P. Evaluating the performance of academic medical education administrators. *Eval Health Prof*. 1999;22:484–496.
- 3 Robbins CJ, Bradley EH, Spicer M. Developing leadership in healthcare administration: A competency assessment tool. *J Healthc Manag*. 2001;46:188–202.
- 4 Jackson E. The 7 reasons why 360 degree feedback programs fail. *Forbes*. August 17, 2012. <http://www.forbes.com/sites/ericjackson/2012/08/17/the-7-reasons-why-360-degree-feedback-programs-fail/#5ebc72044b98>. Accessed February 15, 2017.
- 5 Farey P. Mapping the leader/manager. *Manag Educ Dev*. 1993;24:109–121.

- 6 Palmer M, Hoffmann-Longtin K, Walvoord E, Bogdewic SP, Dankoski ME. A competency-based approach to recruiting, developing, and giving feedback to department chairs. *Acad Med.* 2015;90:425–430.
- 7 Gordon JS, McNew R. Developing the online survey. *Nurs Clin North Am.* 2008;43:605–619, vii.
- 8 Morrel-Samuels P. Getting the truth into workplace surveys. *Harv Bus Rev.* February 2002;5–12.
- 9 Gluck PA. Physician leadership: Essential in creating a culture of safety. *Clin Obstet Gynecol.* 2010;53:473–481.
- 10 Souba WW, Day DV. Leadership values in academic medicine. *Acad Med.* 2006;81:20–26.