



Use of 360° Feedback to Develop Physician Leaders in Orthopaedic Surgery

Paul J. Gregory, PhD¹; David Ring, MD²; Harry Rubash, MD²; and Larry Harmon, PhD¹

Twelve service chiefs participated in 360° feedback surveys and coaching as part of a departmental leadership development activity. Changes in the means of both composite survey scores and individual behavioral item scores over time were evaluated with paired t tests. Agreement between self-rating and rating of others was evaluated with unpaired t tests. There was a nonsignificant change in overall behavioral performance (composite scores) for the physician leaders (n = 12) from baseline [mean (M) = 68.7, standard deviation (SD) = 16.9] to 1-year follow-up (M = 73.1, SD = 11.4), but the performance of four of the leaders with the lowest scores improved substantially. There was a significant improvement in “identifies mistakes respectfully” when comparing baseline to 1-year follow-up. Ten behaviors were identified as improvement opportunities and nine behaviors were identified as behavioral strengths at baseline. Surgeon leaders were in agreement with others’ ratings on 68% of behaviors, underestimated 20% of behaviors, and overestimated 13% of behaviors. (Journal of Surgical Orthopaedic Advances 27(2):85–91, 2018)

Key words: 360° feedback, interpersonal and communication skills, leadership effectiveness, multisource feedback, orthopaedic leaders, PULSE 360, self–other agreement

Effective physician leadership helps meet the challenges of providing coordinated and integrated health care (1, 2). Physician leaders are often promoted to leadership roles on the basis of clinical or research excellence (3) and may be inadequately prepared for the substantial interpersonal demands of their management roles. Surveys of physician leaders have identified communication strategies, listening ability, team building, and conflict resolution as some of the most valued skills. Leaders increase their effectiveness by improving their emotional intelligence (4–6), which can be defined as “effectively understanding oneself and

others, relating well to people, and adapting to and coping with the immediate surroundings to be more successful in dealing with environmental demands” (p.18) (7).

Multisource or 360° feedback surveys can support the development of emotional intelligence and interpersonal skills. Evidence suggests that business managers using ongoing 360° feedback surveys improved more when coaches helped them identify development opportunities compared with those who received no coaching (8). The 360° feedback approach offers several benefits compared with traditional feedback from one’s supervisor alone: (a) a more comprehensive and accurate view of performance; (b) a more thorough and unbiased evaluation than traditional single-source reviews; and (c) a consistent mechanism for assessing and reinforcing behavioral change over time. Coaching helps interpret and accept the feedback, identify strengths and development goals, and create an effective improvement plan (9–12). Turning goals into performance is stronger when a commitment is made and feedback is provided to reinforce progress (13).

To date, physicians have not used 360° feedback and coaching to the same degree as professionals in management and business (14–16). A positive effect of

From ¹PULSE 360 Program/Physicians Development Program, Miami, Florida; ²Massachusetts General Hospital, Boston, Massachusetts. Address correspondence to: Larry Harmon, PhD, Physicians Development Program, 2000 South Dixie Highway, Suite 103, Miami, FL 33133; e-mail: Larry@PdpFlorida.com.

Two of the authors (PG, LH) are employees of the PULSE 360 Program.

Received for publication May 31, 2016; accepted for publication October 5, 2016.

For information on prices and availability of reprints, e-mail reprints@datatrace.com or call 410-494-4994.
1548-825X/18/2702-0085\$22.00/0
DOI: 10.3113/JSOA.2018.0085

feedback and coaching has been demonstrated among residents and non-leader physicians (16–18). Prior research suggests that (a) physicians must be offered a clear and concise explanation of the purpose of the 360° evaluation (19), (b) the process of selecting reviewers ideally should be credible to the feedback recipient (20), (c) the 360° feedback should be used for professional development (20), and (d) the 360° feedback should be debriefed by a feedback coach who can show the physician how he or she scores compared with other physicians, set improvement goals, and encourage the physician review self-study modules regarding emotional intelligence (10). These approaches are likely to help physicians improve their interpersonal and communication skills, professionalism, self-awareness, and teamwork and leadership skills (17, 21).

The leaders of an urban academic orthopaedic surgery department were provided with 360° feedback reports and coaching. This study is a retrospective review of the data collected and the impact of the program. Based on findings in the extant literature, this study addressed four research questions encompassing two a priori hypotheses:

1. Will feedback and coaching lead to improved composite 360° survey scores over time?
2. Is there a significant difference in 360° survey behavioral item scores over time?
3. Are there scoring trends in ratings based on the relative mean scores of the behaviors on the 360° survey (i.e., relative strengths and relative weaknesses)?
4. What behavioral areas are service chiefs' self-ratings in agreement with the ratings provided by others?

Hypothesis 1: Composite 360° scores will show a significant improvement from baseline to follow-up.

Hypothesis 2: Behavioral scores deemed to be improvement opportunities will show significant improvement from baseline to follow-up.

Materials and Methods

Procedure

All 12 service chiefs in the orthopaedic surgery department of a large northeastern academic medical center volunteered to participate in an automated 360° feedback program with coaching (PULSE 360). Each chief selected raters with whom they worked closely (e.g., attendings, inpatient and outpatient staff, management, operating room personnel). The rater list was reviewed and sometimes augmented by the department chief to help control for potential rater selection bias. An e-mail invitation was automatically sent to all raters, emphasizing the anonymity of their survey feedback, highlighting its

developmental purpose, and requesting their candid feedback. An individualized 360° baseline feedback report was generated for each service chief.

A coach from the PULSE 360 program reviewed the results of the survey by telephone with each service chief. The coach and service chief identified a minimum of three leadership "excellence goals" from the survey feedback. Additional coaching was provided to each chief on a voluntary basis. To assess improvement over time, "improvement opportunities" are described in the statistical analyses section below. A follow-up PULSE 360 survey was conducted for each chief approximately 1 year after the baseline survey to evaluate the impact of the feedback, debriefing and goal-setting activity, and additional coaching.

Measures

The PULSE 360 survey consists of 57 items divided into five dimensions: motivating behaviors, discouraging behaviors, motivating impact on others, discouraging impact on others, and open-ended rater comments. The quantitative behavioral item questions ($N = 54$) are rated on 5-point Likert-type scales. The three qualitative open-ended comment questions ask raters to describe the behaviors that they would like the physician to start, stop, and keep doing. A Leadership-Teamwork Index Score (LTI), a proprietary algorithm that combines ratings of all behavioral items on the survey into a single composite survey score ranging from 100 to 100, was computed for each chief. The LTI national average based on a reference data set of physician leaders ($N = 522$) is a mean of 67.5 and standard deviation of 14.8, with a typical observed score range between 0 and 100.

The PULSE 360 survey shows strong internal consistency with Cronbach alphas greater than .85 for all behavioral categories. Additionally, interrater agreement is also high with intraclass correlations above .50 across rater groups (22).

Statistical Analysis

To assess research questions 1 and 2, paired sample t tests were conducted comparing the baseline and follow-up survey results of the chiefs. To assess research questions 3 and 4, the self–other agreement evaluation method recommended by Fleener et al. was used (23). This method categorizes self–other agreement by comparing the mean self-rating score of the sample to the mean others' rating score. Overestimation is identified by a mean self-rating score that is greater than half a standard deviation (of others' rating) above the others' rating score. Underestimation is identified if the self-rating is

TABLE 1 Paired sample *t*-tests comparing change in composite PULSE 360 Leadership-Teamwork Index Score at baseline (BL) vs. follow-up (FO)

PULSE Leadership-Teamwork Index Score	Mean (BL)	SD (BL)	Mean (FO)	SD (FO)	% Difference	<i>t</i>	<i>df</i>	<i>p</i>
All chiefs (<i>N</i> = 12)	68.7	16.9	73.1	11.4	+6.25	1.72	11	.11
Chiefs scoring above national average at baseline (<i>n</i> = 8)	79.8	4.4	78.9	6.9	-1.12	.57	7	.58
Chiefs scoring below national average at baseline (<i>n</i> = 4)	46.6	5.5	61.1	9.2	+31.1%	5.65	3	.011*

* Significant difference at *p* < .02 alpha level.

greater than half a standard deviation below the others' rating score. In-agreement (accurate estimation) is identified by a score that is within the half a standard deviation threshold. An additional type of agreement is in-agreement (unfavorable) where the mean self-rating and mean others' rating meet agreement criteria for a behavioral item that both have rated relatively poorly (a behavioral item score that is more than one-half a standard deviation below the average score of similarly categorized items). In the case of the PULSE 360, all items are categorized into two aggregate behavioral categories: motivating behaviors and impact and discouraging behaviors and impact.

A similar logic was followed to identify behavioral strengths and weaknesses (improvement opportunities). Strengths were identified as any motivating behavior and impact items that had an average others' rating score that was more than one standard deviation above the aggregate motivating behavior and impact item average and was either in-agreement or underestimated. Strengths were also identified as any discouraging behavior and impact items that had a mean others' rating score that was more than one standard deviation below the aggregate discouraging behavior and impact item mean and was either in-agreement or overestimated. Improvement opportunities were identified as any motivating behaviors that had an average others' rating score that was more than one standard deviation below the aggregate motivating behavior item average and was either in-agreement unfavorable or overestimated. Additionally, improvement opportunities

were identified as any discouraging behaviors that had a mean others' rating score that was more than one standard deviation above the aggregate discouraging behavior and impact item mean and was either in-agreement (unfavorable) or underestimated.

Results

There was no significant difference in PULSE 360 survey composite scores between baseline [mean (*M*) = 68.7, standard deviation (SD) = 16.9] and 1-year follow-up (*M* = 73.1, SD = 11.4). In subgroup analysis, the eight chiefs who scored above average (high LTI) at baseline (*M* = 79.8, SD = 4.4) did not improve at follow-up (*M* = 78.9, SD = 6.9), but the four chiefs who scored below average (low LTI) at baseline (*M* = 46.6, SD = 5.5) had significant improvement after a year of coaching and practice (*M* = 61.1, SD = 9.2) (Table 1).

Among the 10 PULSE behavioral items, only "identifies mistakes respectfully" demonstrated significant improvement from baseline (*M* = 3.99, SD = 0.47) to 1-year follow (*M* = 4.11, SD = 0.40) (Table 2). In subgroup analysis, the low-performing chiefs (*n* = 4) also had significant reduction in "creates avoidance" (*M*_{bl} = 2.06, SD_{bl} = 0.42; *M*_{fo} = 1.70, SD_{fo} = 0.32) (Table 3).

The aforementioned analysis plan led to the identification of nine behavioral strengths and 10 behavioral weaknesses (improvement opportunities) based on the others' ratings of the chiefs across all 53 PULSE behavioral items.

TABLE 2 Paired sample *t* tests comparing others' rating mean scores of improvement opportunity PULSE 360 behavioral items at baseline (BL) vs. follow-up (FO) for all chiefs (*N* = 12)

PULSE 360 Behavioral Item	Mean (BL)	SD (BL)	Mean (FO)	SD (FO)	% Difference	<i>t</i>	<i>df</i>	<i>p</i>
1. Adaptive to changes	3.99	0.39	4.02	0.33	+0.7%	-0.49	11	.637
2. Identifies mistakes respectfully	3.99	0.47	4.11	0.40	+3.1%	-2.17	11	.052*
3. Motivates hard work	3.99	0.47	4.09	0.42	+2.6%	-1.58	11	.143
4. Handles difficult team members	3.97	0.45	4.06	0.44	+2.2%	-1.44	11	.177
5. Has social insight	3.87	0.56	3.94	0.46	+1.8%	-0.86	11	.408
6. Criticizes indirectly	1.52	0.34	1.52	0.30	+0.0%	-0.01	11	.993
7. Snaps at others	1.50	0.40	1.44	0.32	-3.8%	0.86	11	.407
8. Creates avoidance	1.44	0.51	1.34	0.33	-7.0%	1.54	11	.152
9. Talks down	1.42	0.42	1.36	0.29	-4.3%	0.98	11	.346
10. Intimidates others	1.41	0.42	1.33	0.32	-5.4%	1.66	11	.125

*Significant difference at *p* < .10 alpha level.

TABLE 3 Paired sample *t* tests comparing others' rating mean scores of improvement opportunity PULSE 360 behavioral items at baseline (BL) vs. follow-up (FO) for low LTI chiefs (*N* = 4)

PULSE 360 Behavioral Item	Mean (BL)	SD (BL)	Mean (FO)	SD (FO)	% Difference	<i>t</i>	<i>df</i>	<i>p</i>
1. Adaptive to changes	3.49	0.19	3.67	0.21	+5.3%	-1.70	3	.188
2. Identifies mistakes respectfully	3.40	0.16	3.66	0.25	+7.9%	-2.34	3	.091*
3. Motivates hard work	3.38	0.15	3.67	0.42	+8.5%	-2.02	3	.136
4. Handles difficult team members	3.43	0.24	3.56	0.27	+3.8%	-1.08	3	.359
5. Has social insight	3.14	0.14	3.42	0.30	+9.0%	-1.68	3	.192
6. Criticizes indirectly	1.82	0.38	1.77	0.29	-2.8%	0.93	3	.419
7. Snaps at others	1.99	0.23	1.77	0.26	-11.2%	1.54	3	.222
8. Creates avoidance	2.06	0.42	1.70	0.32	-17.2%	3.61	3	.037*
9. Talks down	1.88	0.43	1.71	0.23	-9.4%	1.13	3	.340
10. Intimidates others	1.90	0.36	1.71	0.28	-10.0%	1.83	3	.165

*Significant difference at *p* < .10 alpha level.

Behavioral strengths included the following motivating behaviors (higher scored preferred):

1. Shows integrity,
2. Decides effectively,
3. Focused when stressed, and
4. Analyzes before deciding.

Behavioral items also included the following discouraging behaviors (lower score preferred):

5. Yells or swears,
6. Implies retaliation when angry,
7. Makes prejudiced comments,
8. Gets physical when angry, and

TABLE 4 Self-other agreement categorization based on average ratings of orthopaedic service chiefs on PULSE 360° survey motivating behaviors and impact behaviors (higher scores preferred) at baseline

PULSE Behavioral Item	Others' Ratings			Self-Ratings		Self-Other Comparison	
	<i>N</i>	Mean	SD	Mean	SD	Difference	Type
1. Shows integrity	12	4.32	0.31	4.18	0.87	+0.14	In-agreement ^a
2. Decides effectively	12	4.26	0.30	4.27	0.79	-0.01	In-agreement ^a
3. Focused when stressed	12	4.26	0.35	4.27	0.79	-0.01	In-agreement ^a
4. Analyzes before deciding	12	4.25	0.26	4.36	0.81	-0.11	In-agreement ^a
5. Solves problems	12	4.20	0.29	4.18	0.75	+0.02	In-agreement
6. Expresses ideas openly	12	4.20	0.28	4.09	0.70	+0.11	In-agreement
7. Requests respectfully	12	4.18	0.49	4.00	0.77	+0.18	In-agreement
8. Completes on time	12	4.16	0.30	3.82	0.75	+0.34	Underestimation
9. Interacts respectfully	12	4.16	0.48	4.00	0.77	+0.16	In-agreement
10. Timely for commitments	12	4.14	0.28	3.82	0.87	+0.32	Underestimation
11. Helps out	12	4.13	0.39	4.27	0.79	-0.14	In-agreement
12. Listens without interrupting	12	4.11	0.41	3.64	0.81	+0.47	Underestimation
13. Praises others	12	4.10	0.41	3.73	0.65	+0.37	Underestimation
14. Communicates clearly	12	4.08	0.43	3.55	0.52	+0.53	Underestimation
15. Informs others	12	4.08	0.35	3.64	1.03	+0.44	Underestimation
16. Considers suggestions	12	4.06	0.43	4.00	0.63	+0.06	In-agreement
17. Approachable when stressed	12	4.04	0.48	3.91	0.83	+0.13	In-agreement
18. Acknowledges own mistakes	12	4.03	0.39	4.00	0.77	+0.03	In-agreement
19. Motivates best work	12	4.02	0.48	4.09	0.83	-0.07	In-agreement
20. Resolves conflicts	12	4.02	0.45	3.82	0.75	+0.20	In-agreement
21. Adaptive to changes	12	3.99	0.39	3.82	0.75	+0.17	In-agreement ^{b,c}
22. Identifies mistakes respectfully	12	3.99	0.47	3.73	0.90	+0.26	Underestimation ^b
23. Motivates hard work	12	3.99	0.47	4.09	0.83	-0.10	In-agreement ^{b,c}
24. Handles difficult team members	12	3.97	0.45	3.64	0.81	+0.33	Underestimation ^b
25. Has social insight	12	3.87	0.56	3.82	0.75	+0.14	In-agreement ^{b,c}
Average motivating behavior score	12	4.11	0.10	3.95	0.23	+0.16	Underestimation

^aBehavioral strength: relatively high-ranked score by chiefs and others (more than 1 SD above average motivating behavior score mean).

^bImprovement opportunity: relatively low-ranked score by chiefs and others (more than 1 SD below average motivating behavior score mean).

^cIn-agreement (unfavorable) given the relative low ranking of this behavior for both self-ratings and others' ratings.

TABLE 5 Self–other agreement categorization based on the average ratings of orthopaedic service chiefs on PULSE 360° survey discouraging behaviors and impact behaviors (lower score preferred) at baseline

PULSE Behavioral Item	Others'Ratings			Self-Ratings		Self–Other Comparison	
	N	Mean	SD	Mean	SD	Difference	Type
1. Criticizes indirectly	12	1.52	0.34	2.00	0.77	−0.48	Overestimation ^a
2. Snaps at others	12	1.50	0.40	1.91	0.70	−0.41	Overestimation ^a
3. Creates avoidance	12	1.44	0.51	1.55	0.52	−0.11	In-agreement ^{a,b}
4. Talks down	12	1.42	0.42	1.73	0.47	−0.31	Overestimation ^a
5. Intimidates others	12	1.41	0.42	2.18	0.75	−0.77	Overestimation ^a
6. Defensive about suggestions	12	1.39	0.33	1.64	0.50	−0.25	Overestimation
7. Informs only favorites	12	1.39	0.21	1.91	0.54	−0.52	Overestimation
8. Makes negative comments	12	1.38	0.37	1.55	0.52	−0.17	In-agreement
9. Reduces job satisfaction	12	1.35	0.35	1.36	0.50	−0.01	In-agreement
10. Responds late to others	12	1.35	0.20	1.36	0.50	−0.01	In-agreement
11. Responds inappropriately to questions	12	1.33	0.29	1.64	0.81	−0.31	Overestimation
12. Overreacts to mistakes	12	1.32	0.29	1.27	0.47	+0.05	In-agreement
13. Arrogantly demands	12	1.30	0.32	1.36	0.50	−0.06	In-agreement
14. Uses offensive gestures	12	1.30	0.25	1.36	0.50	−0.06	In-agreement
15. Discourages engagement	12	1.30	0.39	1.27	0.47	+0.03	In-agreement
16. Discourages helpfulness	12	1.28	0.28	1.18	0.40	+0.10	In-agreement
17. Discourages questions	12	1.27	0.26	1.36	0.50	−0.09	In-agreement
18. Makes others want to leave	12	1.26	0.26	1.18	0.40	+0.08	In-agreement
19. Avoids responsibilities	12	1.25	0.26	1.09	0.30	+0.16	Underestimation
20. Disrupts team members work	12	1.24	0.26	1.36	0.50	−0.12	In-agreement
21. Interferes with quality work	12	1.24	0.23	1.20	0.42	+0.04	In-agreement
22. Blames others	12	1.24	0.25	1.27	0.47	−0.03	In-agreement
23. Insults others when delays	12	1.21	0.26	1.00	0.00	+0.21	Underestimation
24. Intentionally embarrasses others	12	1.18	0.19	1.00	0.00	+0.18	Underestimation
25. Yells or swears	12	1.09	0.11	1.00	0.00	+0.09	In-agreement ^{c,d}
26. Implies retaliation when angry	12	1.08	0.12	1.00	0.00	+0.08	In-agreement ^{c,d}
27. Makes prejudiced comments	12	1.05	0.06	1.00	0.00	+0.05	In-agreement ^{c,d}
28. Gets physical when angry	12	1.05	0.09	1.10	0.32	−0.05	In-agreement ^{c,d}
29. Makes sexual comments	12	1.04	0.06	1.00	0.00	+0.04	In-agreement ^{c,d}
Average discouraging behavior score	12	1.28	0.13	1.37	0.33	−0.09	Overestimation

^aImprovement opportunity: relatively high-ranked score both by chiefs and others (more than 1 SD above average discouraging behavior score mean).

^bIn-agreement (unfavorable) given the relative low ranking of this behavior for both self-ratings and others' ratings.

^cSelf–other agreement categorization was based on 1SD distance from others' ratings instead of half SD given the very low observed standard deviation in both others' ratings and self-ratings.

^dBehavioral strength: relatively low-ranked score by chiefs and others (more than 1 SD below average discouraging behavior score mean).

9. Makes sexual comments (see Tables 4 and 5).

Behavioral weaknesses or improvement opportunities included the following motivating behaviors:

1. Adaptive to changes,
2. Identifies mistakes respectfully,
3. Motivates hard work,
4. Handles difficult team members, and
5. Has social insight.

They also included the following discouraging behaviors:

6. Criticizes indirectly,
7. Snaps at others,
8. Creates avoidance,
9. Talks down, and

10. Intimidates others (see Tables 4 and 5).

Service chiefs underestimated their behavior in eight motivating behaviors (32%) and three discouraging behaviors (11%). Last, the chiefs overestimated their behavior for no motivating behaviors and seven of 29 (24%) discouraging behaviors (Tables 4 and 5).

Discussion

Feedback surveys with coaching have helped business leaders and managers improve their leadership skills (3, 4, 8, 9, 11). This study applied this approach to service chiefs in an urban academic orthopaedic department. The authors tested for overall improvements over a 1-year period, differences in specific behavioral scores, patterns

in the ratings, and areas where self-ratings matched peer ratings.

There were four major limitations of the current research: the small sample size of service chiefs, the lack of inclusion of other measures of physician leadership (e.g., turnover, morale, profitability, etc.), the coaching and postsurvey feedback activities were not standardized across service chiefs to create uniform support for behavioral improvement, and the year-long time lag in conducting follow-up surveys may be too lengthy because the baseline feedback and goal setting may have begun to lose their impact. This finding reinforces prior research advocating for shorter time frames between iterations of 360° survey feedback as a counter to possible validity threats in the assessment of behavioral change (12). While these findings cannot be generalized to other settings, they do suggest that feedback surveys with coaching can be formative for surgical leaders. Furthermore, casual discussions with both the department and service chiefs about their perceptions of the process and its outcomes were encouraging and positive.

There was partial support for the primary research hypothesis as the authors observed a significant improvement in the subset of service chiefs whose composite survey score was below the national physician leader average at baseline. For the total sample, the change in composite survey score was not significant. As this finding was reviewed more closely, it became apparent that the nonsignificant overall effect may have been driven by the fact that the high LTI (scored above the national average at baseline) service chiefs did not perceive a need to improve their behavior. Prior research has demonstrated that physicians are competency-focused practitioners, which suggests that a service chief who perceives that he or she is already doing a competent job as leader will be less motivated to alter his or her behavior (24–26). The current findings support prior research that physician behavior may best be understood through the theory of reasoned action or theory of planned behavior. Specifically, the low LTI service chiefs were motivated to change their behavior because their lower baseline scores led them to perceive a behavioral deficit and thus to initiate behavioral change. This assertion was upheld in discussion with the service chiefs following their participation in the 360° feedback process and coaching.

There was partial support for hypothesis 2. It was unexpected to find that only one of 10 behavioral items that were identified as improvement opportunities had significantly improved over time. However, on additional evaluation it was discovered that for the low LTI service chiefs there was a significant improvement in the observed mean scores for seven behavioral items overall. Furthermore, those behavioral items were all within the discouraging behavior and impact category (i.e., defensive about

suggestions, overreacts to mistakes, intentionally embarrasses others, discourages questions, discourages helpfulness, creates avoidance, and disrupts team members' work). As with hypothesis 1, the authors believe that the lack of significant differences for the overall sample of service chiefs is being fueled by the limited willingness of the high LTI chiefs to engage in behavior change and also the limited sample size of the group minimizes the power of the statistical analyses. It should be noted that despite the nonsignificant findings, there was a positive trend of improvement for 38 of 54 (70%) behavioral items from baseline to follow-up 1 year later.

The study was able to identify nine behavioral strengths and 10 behavioral weaknesses (improvement opportunities) for the service chiefs. Behavioral strengths in the motivating behaviors and impact items of the survey centered on integrity, decision making, and maintaining focus. Within the discouraging behavior and impact items, the behavioral strengths could be categorized as avoidance of yelling or swearing, prejudiced or inappropriate comments, and aggressive or retaliatory behavior when angry. Behavioral weaknesses in the motivating behavior and impact items included adaptability, motivating others to work hard, handling difficult team members, identifying mistakes in a respectful way, and having insight into how one's behavior affects others. Within the discouraging behavior and impact items, the behavioral weaknesses included criticizing in an indirect manner, snapping or talking down to others, intimidating others, and creating avoidance in team members because of negative behavior. Taken together, these findings support prior research that has demonstrated a competency gap for physician leaders in the realm of communication skills and team motivation (2–4, 15).

Exploring the self–other agreement in the 360° survey ratings of the service chiefs revealed several interesting findings. For instance, service chiefs showed emotional intelligence in that their self-ratings agreed with others' rating for 36 of 54 (67%) behavioral items at baseline. Additionally, the service chiefs underestimated several motivating behavior and impact items and overestimated several discouraging behavior and impact items. This pattern of results indicates that some service chiefs tended to have a humbler or self-deprecating view of their behavior. Furthermore, the service chiefs' pattern of underestimation and overestimation was closely aligned with the behaviors that were identified as behavioral opportunities. However, the service chiefs also identified several behaviors that they seemed to believe were weaknesses that others did not (i.e., time management, listening skills, information dissemination skills, and defensiveness).

Suggestions for future research include studying the generalizability of this 360° survey tool to larger samples

of orthopaedic physician leaders at different settings; the impact of additional interventions on improvement such as standardized education, increased frequency of goal reminders, periodic feedback on goal achievement, and sharing feedback and goals with team members; the impact of finances on improving behavior as assessed by 360° feedback (i.e., “pay-for-professionalism”) (22); the effectiveness of a 360° feedback approach for nonphysician management such as department administrators and nurse managers; the potential relationship between 360° ratings with other physician leader metrics described above; and the applicability of 360° feedback to develop physician leaders in other specialties.

Based on the feedback and findings of this project, six recommendations are proposed for implementing a successful 360° survey-based program for orthopaedic surgery physician leaders:

1. Create physician leader and rater engagement by emphasizing the leadership development purpose of the process;
2. Balance the rater selection process by inviting both participants as well as their leaders to select from a large pool of potential raters to help ensure both comprehensive and diverse feedback perspectives;
3. Provide anonymity protections for raters’ anecdotal comments by grouping them into behavioral themes as well as to help leaders to focus more on feedback content rather than its source;
4. Use a trained mentor or coach to review results, address any pushback, and set leadership excellence goals;
5. Provide ongoing coaching and educational modules for leaders with outlying scores or comment themes; and
6. Conduct a short follow-up survey of the same raters within 3 to 6 months of the baseline to provide feedback about any behavioral changes to reinforce improvement.

The authors believe that the proposed recommendations will enable orthopaedic surgery and other departments to create a formative and beneficial leadership development process that uses tools and procedures that encourage feedback that is more comprehensive, objective, and candid than the traditional feedback systems employed in health care.

References

1. Stoller, J. K. Developing physician-leaders: key competencies and available programs. *J. Health Adm. Educ.* 25:307–328, 2008.
2. Stoller, J. K. Developing physician-leaders: a call to action. *J. Gen. Intern. Med.* 24(7):876–878, 2009.
3. McAlearney, A. S., Fisher, D., Heiser, K., et al. Developing effective physician leaders: changing cultures and transforming organizations. *Hosp. Top.* 83(2):11–18, 2004.
4. Mintz, L. J., Stoller, J. K. A systematic review of physician leadership and emotional intelligence. *J. Grad. Med. Educ.* 6(1):21–31, 2014.
5. Sosik, J. J., and Megerian, L. E. Understanding leader emotional intelligence and performance the role of self-other agreement on transformational leadership perceptions. *Group Organ. Manage.* 24(3):367–390, 1999.
6. Bratton, V. K., Dodd, N. G., Brown, F. W. The impact of emotional intelligence on accuracy of self-awareness and leadership performance. *Lead. Organ. Dev. J.* 32(2):127–149, 2011.
7. Bar-On, R. M. The Bar-On model of emotional-social intelligence (ESI). *Psicothema.* 18(1):13–25, 2006.
8. Smither, J. W., London, M., Flautt, R., et al. Can working with an executive coach improve multisource feedback ratings over time? A quasi-experimental field study. *Pers. Psychol.* 56(1):23–44, 2003.
9. Luthans, F., Peterson, S. J. 360-Degree feedback with systematic coaching: empirical analysis suggests a winning combination. *Hum. Resource Manage.* 42(3):243–256, 2003.
10. Hammerly, M., Harmon, L., Schwaitzberg, S. Good to great: using 360-degree feedback to improve physician emotional intelligence. *J. Healthcare Manag.* 59(5):354–365, 2014.
11. Alexander, D. M. How do 360 degree performance reviews affect employee attitudes, effectiveness and performance. Paper presented at Schmidt Labor Research Center Seminar Research Series, 2006.
12. Alimo-Metcalfe, B. 360 degree feedback and leadership development. *Int. J. Select. Assess.* 6(1):35–44, 1998.
13. Locke, E. A., Latham, G. P. Building a practically useful theory of goal setting and task motivation. *Am. Psychol.* 57:705–717, 2002.
14. Fleenor, J. W., Taylor, S., Chappelow, C. *Leveraging the Impact of 360-Degree Feedback*, John Wiley & Sons, Hoboken, NJ, 2008.
15. Wolper, L. F. *Physician Practice Management: Essential Operational and Financial Knowledge*, Jones & Bartlett Publishers, Boston, MA, 2012.
16. Overeem, K., Wollersheim, H., Driessen, E., et al. Doctors’ perceptions of why 360-degree feedback does (not) work: a qualitative study. *Med. Educ.* 43(9):874–882, 2009.
17. Violato, C., Lockyer, J., Fidler, H. Multisource feedback: a method of assessing surgical practice. *BMJ* 326(7388):546–548, 2003.
18. Higgins, R. S., Bridges, J., Burke, J. M., et al. Implementing the ACGME general competencies in a cardiothoracic surgery residency program using 360-degree feedback. *Ann. Thorac. Surg.* 77(1):12–17, 2004.
19. Sargeant, J., Mann, K., Ferrier, S. Exploring family physicians’ reactions to multisource feedback: perceptions of credibility and usefulness. *Med. Educ.* 39(5):497–504, 2005.
20. Sargeant, J. M., Mann, K. V., Ferrier, S. N., et al. Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: a pilot study. *Acad. Med.* 78(10):S42–S44, 2003.
21. Harmon, L. Lapenta, S. Managing difficult and disruptive physicians. In *The Business of Healthcare, Vol 1: Practice Management*, pp. 75–91, edited by K. Cohn and D. Hough, Praeger, New York, 2008.
22. Hageman, M. G., Ring, D. C., Gregory, P. J. Do 360-degree feedback survey results relate to patient satisfaction measures? *Clin. Orthop. Relat. Res.* 473(5):1590–1597, 2015.
23. Fleenor, J. W., Smither, J. W., Atwater, L. E., et al. Self–other rating agreement in leadership: a review. *Lead. Q.* 21(6):1055–1034, 2010.
24. Perkins, M. B., Jensen, P. S., Jaccard, J., et al. Applying theory-driven approaches to understanding and modifying clinicians’ behavior: what do we know? *Psychiatr. Serv.* 58(3):342–348, 2007.
25. Grol, R. Changing physicians’ competence and performance: finding the balance between the individual and the organization. *J. Cont. Educ. Health Prof.* 22(4):244–251, 2002.
26. Smith, W. R. Evidence for the effectiveness of techniques to change physician behavior. *Chest J.* 118(2 Suppl.):8S–17S, 2000.