Quality and Patient Safety

Learning and Improving in Quality Improvement Collaboratives: Which Collaborative Features Do Participants Value Most?

Ingrid M. Nembhard

Objective. To understand participants' views on the relative helpfulness of various features of collaboratives, why each feature was helpful and which features the most successful participants viewed as most central to their success.

Data Sources. Primary data collected from 53 teams in four 2004–2005 Institute for Healthcare Improvement (IHI) Breakthrough Series collaboratives; secondary data from IHI and demographic sources.

Study Design. Cross-sectional analyses were conducted to assess participants' views of 12 features, and the relationship between their views and performance improvement.

Data Collection Methods. Participants' views on features were obtained via self-administered surveys and semi-structured telephone interviews. Performance improvement data were obtained from IHI and demographic data from secondary sources.

Principal Findings. Participants viewed six features as most helpful for advancing their improvement efforts overall and knowledge acquisition in particular: collaborative faculty, solicitation of their staff's ideas, change package, Plan-Do-Study-Act cycles, Learning Session interactions, and collaborative extranet. These features also provided participants with motivation, social support, and project management skills. Features enabling interorganizational learning were rated higher by teams whose organizations improved significantly than by other teams.

Conclusions. Findings identify features of collaborative design and implementation that participants view as most helpful and highlight the importance of interorganizational features, at least for those organizations that most improve.

Key Words. Quality improvement collaboratives, organizational learning

Quality improvement collaboratives are an increasingly common strategy for implementing evidence-based practices in health care. Most collaboratives are modeled after the Breakthrough Series (BTS) model, developed by the

Institute for Healthcare Improvement (IHI). In the BTS collaborative model, IHI identifies a topic for which a large gap exists between actual and recommended care, assembles expert faculty to develop a "change package" of evidence-based practices to close the gap, and invites organizations to join the effort. Multidisciplinary teams from participating organizations then work together on the specified topic for approximately a year. Teams attend a series of multiday meetings, known as "Learning Sessions," where they learn improvement techniques from experts and share their experiences implementing new practices with one another. Between Learning Sessions, teams implement changes in their own organizations using Plan-Do-Study-Act cycles in which they investigate quality problems, develop and implement small-scale changes, measure the effects, and make various changes for improvement. Teams learn from one another by participating in monthly conference calls, team-to-team telephone calls, listserv discussions, site visits to other organizations, monthly exchange of written reports detailing improvement activities, and monthly posting of performance data to the collaborative extranet. After the collaborative ends, teams summarize their results and lessons learned, and present them to nonparticipating organizations at conferences (Kilo 1998, 1999; IHI 2003).

Previous studies of the impact of collaboratives on organizational performance have mixed results, with some demonstrating substantial impact on improvement (Horbar et al. 2001; Dellinger et al. 2005; Howard et al. 2007) and others finding no significant effect (Landon et al. 2004; Homer et al. 2005). The vast majority of these studies have examined collaboratives as a single intervention as opposed to the diverse set of features that comprise most collaborative models. A smaller group of studies has described the various features of collaboratives (e.g., Kilo 1998; Wilson, Berwick, and Cleary 2003; Ayers et al. 2005), but none has evaluated the influence of these various features on performance, particularly from the perspective of collaborative participants. Research by Fremont et al. (2006) and Leape et al. (2006) are notable exceptions. However, the former examined only three features of collaboratives, and the latter included a restricted sample of a single-state collaborative. Moreover, neither study empirically tested the relative helpfulness of various features of the collaboratives. Thus, we know little about which features are viewed as most helpful, why they are helpful, and which features

Address correspondence to Ingrid M. Nembhard, Ph.D., M.S., Assistant Professor of Public Health and Management, Yale University School of Medicine and Yale University School of Management, 60 College Street, New Haven, CT 06520-8034; e-mail: ingrid.nembhard@ yale.edu.

are rated most highly by participants whose organizations are most successful in improving performance. This knowledge is central for researchers, practitioners, and policy makers seeking to use collaboratives effectively to improve quality of care.

Therefore, we sought to examine participants' views on the relative helpfulness of various features of collaboratives, why each feature was helpful, and which features the most successful participants viewed as most central to their success. We used quantitative and qualitative data from participants (53 teams) in four 2004–2005 BTS collaboratives. Findings from this study inform ongoing discussions about the most effective activities for promoting quality improvement. They also inform debates about whether the collaborative model can be modified to make it accessible to resource-constrained organizations without sacrificing the most valued features. For example, some have questioned whether multiday, face-to-face Learning Sessions, which require additional financial and nonfinancial investments are fundamental or can be replaced by 1-day or virtual sessions with equal success (Fremont et al. 2006). The findings from this study offer insight into which collaborative features should be maintained because of their value for participants.

METHODS

Study Design and Sample

We conducted a cross-sectional study of participants' views of 12 design and implementation features of collaboratives, using survey, archival, and interview data. Participants were teams from organizations in the United States and Canada who had participated in one of four BTS collaboratives sponsored by IHI in 2004–2005. The topics of the collaboratives were: Improving Access and Efficiency in Primary Care (ACCESS), Reducing Complications from Ventilators and Central Lines in the ICU (ICU), Reducing High Hazard Adverse Drug Events (ADE), and Reducing Surgical Site Infections (SSI).

Of the 78 teams contacted for the survey, 67 (86 percent) teams had at least one person return a survey. Of these 67 teams, we eliminated teams with less than three members responding, consistent with recommendations for improving reliable measurement of group experiences (Van der Vegt and Bunderson 2005; Marsden et al. 2006). The resulting sample consisted of 53 teams and a total of 217 member respondents, for an effective response rate of 68 percent (53/78 teams). On average, there were 13 teams per collaborative and four respondents per team.

Data Collection

Survey Data. We invited team members in all four collaboratives to participate in our survey about their experiences. A letter, distributed via collaborative listserv, explained that the survey would be distributed to all team members present at the final Learning Session of the collaborative, and offered the opportunity for online or paper completion of the survey to those unable to attend the meeting. Following the Learning Session, five surveys with self-addressed stamped envelopes were mailed to the key contact for each team that did not attend. A reminder e-mail to complete the survey with "thank you" to those who had participated was then sent to all team members within 2 weeks of the Learning Session.

Archival Data. We obtained performance improvement data for each team's home organization from IHI records. We obtained demographic data on each organization from the American Hospital Association Directory, Profiles of U.S. Hospitals, and organizations' websites.

Interview Data. We conducted semi-structured telephone interviews with nine randomly selected team leaders (two team leaders per collaborative for three collaboratives, and three leaders for the fourth) about 1 month after the end of the collaborative. The interviewer was blind to the survey and performance data. During the interviews, which lasted from 35 minutes to 1 hour, interviewees were asked to reflect on their experiences during the collaborative. For example, we asked: "If you were advising a team that was joining a collaborative for the first time, what advice would you give them? Which collaborative activities would you recommend they participate in?" Interviews were audio-taped and transcribed to create a qualitative data set.

Measures

Overall Feature Helpfulness. We assessed the overall helpfulness of 12 features of collaborative design and implementation (Table 1). These features were identified through review of the literature on BTS collaboratives (Kilo 1998, 1999; IHI 2003; Wilson, Berwick, and Cleary 2003), personal observations and conversations with IHI faculty. We asked survey respondents to rate "how helpful your team found [each feature] for carrying out improvement projects during the collaborative" using a five-point scale ($1 = $ no help, $2 = $ little help, $3 = $ moderate help, $4 = $ great help, $5 = $ very great help). We

Table 1:   Collaborative Features

Feature	Description
Change package	A toolkit of evidence-based practices and implementation strategies that includes an explanation of the rationale for each recommended practice, appropriate progress measures, data collection techniques, tools (e.g., sample forms and policy statements) and reference materials
Collaborative faculty	Experts in the topic area that teach teams improvement techniques and provide additional guidance to teams as needed
Learning Sessions interactions	Formal and informal interactions with other teams during the multiday, face-to-face meetings; formal interactions involved teams sharing their experiences implementing new practices
Monthly conference calls	Conference calls organized by the collaborative faculty, typically with a planned discussion topic
Team-initiated phone calls	Telephone calls by teams to other teams in the collaborative
Listserv discussions	E-mail communications via listserv, typically involving information exchange, and asking and answering questions
Site visits	Visits by teams to other organizations in the collaborative to observe and discuss practice implementation
Monthly report exchange	Written progress report by teams in prescribed template that documents past month's activities and self-assessment of progress
Collaborative extranet	Password-protected Internet site where teams could post their performance data and information; only viewable by participants
Plan-Do-Study-Act (PDSA) cycles	Rapid cycle change methodology, developed by W. Edwards Deming, in which teams investigate quality problems, develop (Plan) and implement small-scale changes (Do), measure the effects (Study), and make changes until satisfied with outcomes (Act)
Solicitation of staff ideas	Team's solicitation of ideas and feedback from staff in their home organization
Literature reviews	Retrieval and review of published information on a defined topic

then averaged team members' individual ratings for each feature to create a team-level rating because respondents were asked to report the team experience and their responses were relatively homogenous at the team level, as indicated by median interrater agreement scores (James, Demaree, and Wolf 1984) >.70 for all features (Glick 1985).

Feature Helpfulness for Knowledge Acquisition. After offering an overall assessment of each feature, survey respondents evaluated feature helpfulness for acquiring two types of knowledge scholars theorize are obtained via collaboratives (Kilo 1999): general knowledge (know-what) and

implementation knowledge (know-how). Specifically, respondents assessed, "how helpful your team found [each feature] for generating each type of information: general knowledge about potentially better practices and actual strategies for practice implementation," using a five-point scale (1 = no help, 2 = little help, 3 = moderate help, 4 = great help, 5 = very great help). Individual scores were aggregated to the team level.

Performance Improvement. Each organization's improvement was assessed by the IHI Director who was assigned to each collaborative using organization-submitted data for specified metrics (e.g., for the SSI collaborative, the number of surgical cases between surgical site infections). IHI Directors classified the level of improvement as no improvement, modest improvement, improvement, or significant improvement using defined criteria. For example, to be classified as "significant improvement," an organization had to have implemented the majority of the recommended practices, demonstrated evidence of breakthrough improvement in specified outcome measures (e.g., SSI breakthrough goal: double the number of surgical cases between surgical site infections), been at least 50 percent toward accomplishing its goals, and established plans for spreading improved practices throughout the organization. We coded teams whose organizations experienced significant improvement, the most successful, as 1; all other teams were coded as 0.

Covariates. Team size, collaborative topic, and prior collaborative experience were included as covariates in analyses involving performance improvement because prior research shows they influence performance (Simonin 1997; Shortell et al. 2004; Landon et al. 2007). However, they were excluded from other analyses because preliminary regression models showed they had no significant association to helpfulness ratings ($p > .05$).

Data Analysis

We used standard frequency analyses to examine the characteristics of sample participants. To assess participants' views on the relative helpfulness of various features, we averaged their ratings of helpfulness for each feature. This led to the creation of a mean helpfulness score for each feature. We used one-sample *t*-tests to determine whether these scores differed significantly from a score of 3.5 ($p < .05$). We used 3.5 as our reference value because a score higher than

3.5 indicated that a feature was more than moderately helpful for improvement work; it was a great help.[1] We wanted to identify these great features. We calculated Cohen's (1977) effect size index (d) for significant test results to examine the magnitude of effects.

To understand the purpose each feature served, we compared each feature's helpfulness for acquiring general knowledge versus implementation knowledge using paired *t*-tests. Test results enabled us to categorize each feature as primarily a means to acquiring general knowledge, implementation knowledge, or both equally. We then used one-sample *t*-tests to assess whether the mean helpfulness score for acquiring the primary (higher) knowledge type was significantly higher than 3.5 ($p < .05$). A score higher than 3.5 indicated that a feature was more than moderately helpful for acquiring new knowledge; it was a great help.

To examine differences in ratings of feature helpfulness between the more and less successful teams, we estimated a separate general linear model for each feature. The models evaluated the association between making significant improvement as defined by the IHI Director and helpfulness ratings for the selected feature. Models included the three covariates.

In our analyses, sample sizes for features varied because the number of teams with experience with each feature varied. Nevertheless, most features were evaluated by more than 50 teams.

To improve our understanding of the quantitative findings, we analyzed the qualitative data from our interviews. The interviewer—consistent with the anthropological viewpoint that a single interviewer–analyst is preferred to capitalize on the interviewer's knowledge of the research instrument (Morse 1994; Morse and Richards 2002; Janesick 2003)—transcribed the interview tapes, entered the transcripts into *Atlas.ti*™ data analysis software, and coded the transcripts, using the constant comparative method (Strauss and Corbin 1998), for common themes concerning the collaborative features, evaluation criteria, and feature purpose.

RESULTS

Participants' Characteristics and Performance Improvement

In our sample of 53 teams, 38 percent of teams were from the Northeast U.S., 81 percent from urban settings, 64 percent from teaching institutions, 36 percent from health systems, and 98 percent from not-for-profit institutions. Fifty-one percent of teams ($N = 27$) were from organizations that were classified by the collaborative director as having experienced "significant improvement,"

38 percent (*N* = 20) as "improvement" and 11 percent (*N* = 6) as "modest improvement." Differences between the 53 teams in the study sample and the 25 nonstudy teams in terms of geography, setting (urban versus rural), teaching status, health system membership, ownership status, organizational size, and performance improvement were not statistically significant ($p > .05$).

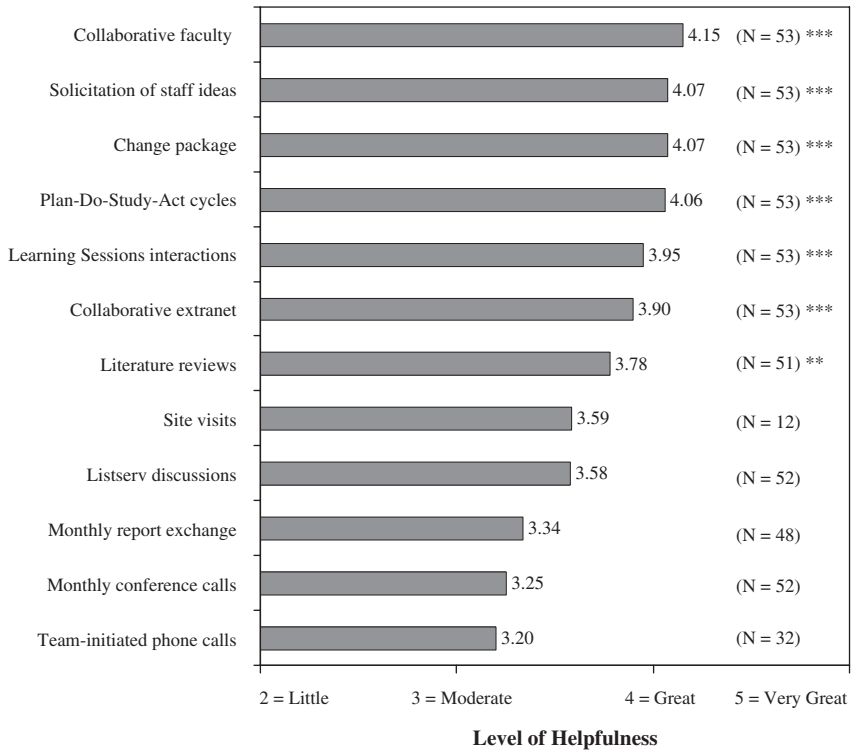## Overall Feature Helpfulness

Several features—collaborative faculty, solicitation of staff ideas, change package, PDSA cycles, Learning Sessions interactions, collaborative extranet, and literature reviews—received helpfulness ratings that were significantly better than 3.5 ($p < .01$), suggesting that these features were more than moderately helpful to teams as they carried out their improvement work (Figure 1). They were a great help. The mean (median) Cohen's (1977) effect size for these features was 0.87 (0.95), a large effect size. The five remaining features received scores that were not statistically different from 3.5 or were significantly less than 3.5 ($p < .05$), suggesting that these features were at most moderately helpful. No scores were adjusted for demographic or covariate characteristics because preliminary tests showed no significant association between these characteristics and helpfulness ratings for any of the 12 features ($p > .05$).

Interviewees' comments suggested several reasons for the tempered evaluations of the moderately helpful features, whose scores had a larger mean standard deviation (0.86) than great help features did (0.58). Complimentary remarks were accompanied by caveats, criticisms, and comments on circumstantial realities, more so than for great features. For example, ICU team leader 133 said:

> The listserv is great, but people just feel intimidated to use it because they don't want to ask a question that people will think is stupid. Any time we had a question, I asked my team, 'Is this something we want to post?' but none of [them] felt comfortable to post it.

This team leader would post anyway. Her story exemplifies one barrier to realizing the full benefit of these features: reputation concerns. Some of these features required publicly revealing ignorance, an uncomfortable act that some elected not to experience. This would have been the case for the ICU team described just above if not for its leader. In the absence of leaders' persistence or other countervailing forces (e.g., team learning orientation), fear of embarrassment limited participation in the more interactive features for some teams, and thus minimized these features' potential helpfulness.

Figure 1: Participants' Ratings of Overall Feature Helpfulness[§]

Feature	Rating	
Collaborative faculty	4.15	(N = 53) ***
Solicitation of staff ideas	4.07	(N = 53) ***
Change package	4.07	(N = 53) ***
Plan-Do-Study-Act cycles	4.06	(N = 53) ***
Learning Sessions interactions	3.95	(N = 53) ***
Collaborative extranet	3.90	(N = 53) ***
Literature reviews	3.78	(N = 51) **
Site visits	3.59	(N = 12)
Listserv discussions	3.58	(N = 52)
Monthly report exchange	3.34	(N = 48)
Monthly conference calls	3.25	(N = 52)
Team-initiated phone calls	3.20	(N = 32)

2 = Little   3 = Moderate   4 = Great   5 = Very Great

Level of Helpfulness

Notes: [§]Helpfulness ratings were compared using one-sample t-tests to a score of 3.5, the upper limit score for a moderately helpful feature. A rating significantly higher than 3.5 indicates a feature was more than moderately helpful to teams as they carried out their improvement work. They were a great help.

** = $p < .01$

*** = $p < .001$

Cohen's (1977) effect size index (d) for statistically significant features is as follows: 0.98 for collaborative faculty, 1.22 for solicitation of staff ideas, 1.00 for change package, 0.95 for PDSA cycles, 0.80 for Learning Session interactions, 0.71 for collaborative extranet, and 0.43 for literature reviews. An effect size of 0.20 suggests a small effect, 0.50 a medium effect, and 0.80 a large effect.

In most cases however, interviewees reported that their teams felt comfortable seeking help. Their disappointment in moderately helpful features stemmed from the quantity of the feedback received. Responses to listserv postings were not always received. The same was true for questions posed in monthly conference calls. Interviewees guessed that nonresponse signaled others' busyness or failure to locate an answer themselves. Whatever the reason, they were not helped by the absence of response, which tempered enthusiasm for these features.

Another factor that weighed into helpfulness evaluations was the relevance of the knowledge shared. Although interviewees mostly praised features for providing access to needed knowledge, some noted that variability in the relevance of the knowledge obtained via moderately helpful features lowered their assessment of these features. For example, SSI team leader 170 remarked:

> The conference calls could be iffy. Some were good, some were not. Some were like 'We're wasting our time here.' Others were, 'Oh, we need to get that.' It depended on the topic more than anything else.

Participants felt some call topics had broad appeal, containing information that all teams found helpful, while other calls benefited a limited number of teams. Consequently, several teams considered conference calls less than great. Great calls would have incorporated more time for unstructured conversation, allowing all teams to raise their most pressing issues. Without enough of this time, some teams felt like ACCESS team leader 112 described, "The monthly calls were okay. If there was more time for Q&A from each other that would have been more helpful."

Another factor that influenced helpfulness ratings was the participant role. For example, in some site visits, the host served primarily as a teacher due to its success, while the visitor served as a student. These visits were helpful for the visitor, but not always for the host. Thus, one team leader whose organization hosted several site visits said that her organization planned to reduce the number of future visits because visits distracted from its own improvement work.

Together, interviewees' comments about moderately helpful features suggested that three factors led to lower scores: participant role for site visits; limited use for listserv and phone calls (conference and team-initiated); and limited quantity and relevance of knowledge gained for conference calls, listserv, and monthly reports. Of the factors, the last appeared most influential.

Feature Helpfulness for Acquiring Knowledge

Six features were viewed as a great help for advancing either general knowledge or implementation knowledge (Table 2). Three features helped participants acquire general knowledge (collaborative faculty, change package, and collaborative extranet), two helped with implementation knowledge (PDSA

Table 2: Acquiring General and Implementation Knowledge: Primary Contributing Features and Their Level of Helpfulness

Level of Helpfulness	Type of Knowledge		
	General Knowledge	*Implementation Knowledge*	*General* and *Implementation Knowledge*
Great	Collaborative faculty ($M = 4.08$)	PDSA cycles ($M = 4.00$)	Learning Session interactions ($M = 3.95$)
	Change package ($M = 4.05$)	Solicitation of staff ideas ($M = 4.00$)	
	Collaborative extranet ($M = 3.75$)		
Moderate	Literature reviews ($M = 3.65$)	Team-initiated phone calls ($M = 3.04$)	Listserv discussions ($M = 3.51$)
			Site visits ($M = 3.50$)
			Monthly conference calls ($M = 3.25$)
			Monthly report exchange ($M = 3.02$)

Note: Reported means are the mean helpfulness scores for delivering that particular knowledge. In the case of features that equally delivered both general and implementation knowledge, meaning there was no significant difference in the mean helpfulness scores between the knowledge types, the higher of the mean scores is reported.

Illustrative Comments

Faculty and Change Package: They [IHI] have a lot of good strategies and theories and tools, but making it actually work in an organization is very different . . . We know this is the right thing to do, but how are we actually going to make it happen? How do you actually get the staff to do this? How do you actually get the physicians to do it? What are the key things you should try to do?

Staff ideas: We learned of the administrative walkarounds and safety huddles at the Learning Session and through the collaborative, but as we started to do them we learned a lot from front-line staff here in our organization. —ADE team leader 154

PDSA cycles and staff ideas: We think one of the ways we tempered our frustration with implementing new

Learning Sessions: We think the best thing is going to the Learning Sessions and actually talking to people that have actually done it because they actually know how you actually do it. We get the expertise from them, and then get it internally. —ICU team leader 131

Learning Sessions: When we got together for the Learning Sessions and really talked face-to-face, that's what

Table 2:   *Continued*

	Illustrative Comments	
Who are the people who should say the message? You know—the real specifics. That's really important, and sometimes we don't get that detail. —ICU team leader 131 *Faculty and Change Package:* Most of what IHI recommends you do is evidence-based. One of the key foundational components is that it's stuff that's been proven to be good practice. It's figuring out how to deliver those best practices in a reliable way. It's the how-to, not the what, that's the challenge. —ICU team leader 139	practices was trusting the process and believing in the process. We trusted that if you just keep doing these small tests of change, and constantly get feedback from the staff about why it isn't working, then constantly tweak your process, you get to the right process. Trusting that that's the right thing to do and doing it helped us. —ICU team leader 139	helped us the most. That was really the best. —SSI team leader 169 *Listserv:* Our group probably used the listserv every week until the last couple of months, and then maybe we started using it a little less. Usually what we were using it for was some sort of advice about how to get something up and running. —ICU team leader 139 *Listserv:* Online, when someone would ask a question like, "How do you do this?" and then you get responses, that was some of the most useful, practical information other than the first Learning Session. —ACCESS team leader 112

cycles and solicitation of staff ideas) and one helped equally with both (Learning Session interactions). Notably, with the exception of Learning Session interactions, features that equally offered both knowledge types were regarded as only moderately helpful for acquiring both. Also noteworthy, for the majority of features, helpfulness scores for knowledge acquisition fell below overall helpfulness scores, suggesting that participants valued features for more than their knowledge benefits.

Interviewees spoke about three additional benefits of features. First, they provided access to inspiration and motivation. Learning what others achieved motivated participants to persist with their improvement efforts. ADE team leader 146 explained:

> When we went to the Learning Session and saw hospitals that were going months without a ventilator-associated pneumonia, it was like 'Oh my God. We can do better.' So, we set our own targets higher.

Second, use of the more interactive features of the model provided access to social support, which made the challenge of improvement work more bearable. As ICU team leader 131 explained:

> It was nice to know that people were dealing with the same issues back at their hospital. So, it's not that it provided answers. It provided a sense of what everybody else is dealing with too. That makes it a little easier—the social connection. You're trying to sustain the change. To hear that it's difficult, it just helps.

Third, interviewees felt the project management structure imposed by the collaborative, particularly monthly reporting, was helpful. Many expressed the same sentiment as SSI team leader 169:

> You know monthly you've got to send in all of this stuff [report and data to the extranet]. That part is burdensome, but on the other hand, it gets you where you need to be. Because you must do this by this date, you get a faster reward.

In sum, participants believed that improvement resulted not only from knowledge gains, but also because features provided them with motivation, social support, and project management skills.

Linking Perceived Feature Helpfulness to Performance Improvement

Figure 2 shows the similarities and differences in overall helpfulness ratings between teams whose organizations made significant improvement and teams whose organizations made less improvement—for all features except the two

Figure 2:   Comparison of Helpfulness Ratings by Participants that Improved Significantly and Participants that Did Not
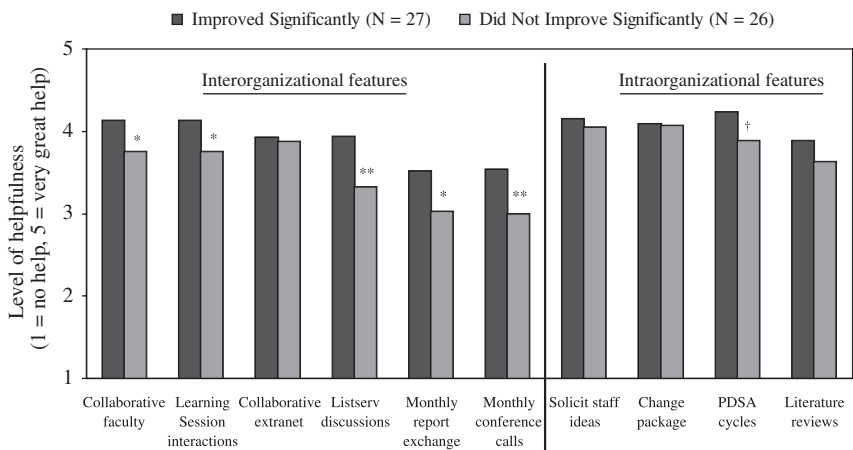
$\dagger = p < .10$; $* = p < .05$; $** = p < .01$

few teams utilized (site visits and team-initiated phone calls). Features to the right of the vertical line—solicitation of staff ideas, change package, PDSA cycles, and literature reviews—were rated the same by both groups ($p > .05$). All were a great help to improvement work. Notably, these features did not involve knowledge-sharing with other teams. Combining this observation with the helpfulness result indicates both groups believed *intra*organizational learning contributed greatly to their improvement.

The two groups disagreed on the helpfulness of features enabling *inter*organizational learning, such as Learning Sessions, listserv discussions, and monthly conference calls (features to the left of the vertical line in Figure 2). Teams from organizations that made significant improvement rated these features, with the exception of the collaborative extranet, significantly higher in helpfulness than teams from organizations that made more modest improvement ($p < .05$). This difference in ratings may in part explain why helpfulness scores for most of these features were tempered. Scores appear to reflect the average of two groups with divergent views.

DISCUSSION

Our results showed that participants valued all of the 12 features studied. However, they regarded six features as most valuable for both advancing improvement efforts overall and (general and implementation) knowledge acquisition in particular. These features were: collaborative faculty, solicitation of staff ideas, change package, PDSA cycles, Learning Session interactions, and collaborative extranet. These features also helped participants maintain their motivation, access social support, and improve their project management skills.

Most notably, our results revealed that features that enable interorganizational learning are viewed as significantly more helpful by participants from organizations that experienced significant improvement than by participants from organizations that did not. This finding suggests a potential explanation for the mixed results of research on collaborative effectiveness. The use of interorganizational features, and interorganizational learning in turn, may mediate the impact of collaborative membership on organizational performance. This potential mediating effect suggests the importance of not only assessing "Do collaboratives work?" but also assessing "What [about collaboratives] works for whom in what circumstances?" (Pawson and Tilley 1997). Our results suggest that collaboratives may work most for participants who

capitalize on their interorganizational features in addition to their intraorganizational features. This hypothesis is consistent with research in other industries showing performance benefits of combining inter- and intraorganizational activities (e.g., Ancona and Caldwell 1992).

Our findings raise the question: why did teams from organizations that experienced less improvement view features that enable interorganizational learning as less helpful? One possibility is that less successful teams were unable to capitalize on the benefits of these features due to internal or external constraints such as lack of management support, an unsupportive organizational culture, or poor team functioning, factors other studies show influence performance improvement (Mills and Weeks 2004; Shortell et al. 2004; Bradley et al. 2006). A second possibility is that these teams suffered a misattribution error. Attribution theorists have shown that individuals attribute their poor performance to situational factors outside of their domain (Jones and Nisbett 1971; Gilbert and Malone 1995). Thus, poorer-performing teams may have negatively attributed their performance to features that involved others. A third possibility is that these features are inherently less helpful than better performers claim. Better performers may rate all features highly, overestimating the effect of some, because they are overly enthusiastic about the collaborative experience. Unfortunately, we cannot identify the true reason with our data.

We studied participants in four collaboratives that varied in clinical focus, and found no significant difference in helpfulness ratings across collaboratives. However, the helpfulness of features may vary for a different set of collaborative topics. For instance, features that facilitate interorganizational learning may be less helpful than features that facilitate intraorganizational learning when the practices recommended by the collaborative require substantial adaptation to fit the organizational context. The possibility that feature helpfulness is contingent upon new practice characteristics is consistent with research showing that hospital units in which improvement teams used learning activities that facilitated the adaptation of context dependent practices experienced greater implementation success (Tucker, Nembhard, and Edmondson 2007).

Our survey results suggest that teams do not find site visits extremely valuable. However, we caution against concluding that site visits are not a great help to participants. Our small sample size for users, who rated this feature ($N = 12$), may have limited our power to detect significant differences. Moreover, our qualitative data suggest that site visits can be a great help for two reasons. First, they afford the visitor an opportunity to observe recom-

mended practices in operation. Research on best practice transfer suggest that such observation is beneficial for practice implementation because many new practices in health care have a large tacit component that is not easily described or codified (Berta and Baker 2004). Second, site visits grant visitor and host greater opportunity to interact and share ideas and experiences related to the topic area. The two organizations that participated in site visits in Fremont et al.'s (2006) qualitative study echoed this view, supporting the high value of this feature for collaborative participants.

A central question for designers and implementers of collaboratives is whether and how to modify the model to increase its effectiveness for improving quality of care. Our findings imply that modifications that reduce the emphasis on the six great features are ill-advised because they would diminish the value of collaboratives from the participant perspective. Decreasing Learning Sessions or replacing them with virtual sessions, for example, would reduce the opportunities for participant interactions, a most helpful feature of collaboratives. Whether increasing the emphasis on the six great features would yield additional benefits or whether there are diminishing returns due to resource constraints or information saturation is a question that requires further study.

Although our study's findings are informative, they should be considered in light of the methodological limitations. First, our effective response rate of 68 percent is less than ideal; however, this response rate is comparable to other studies of collaboratives (Landon et al. 2004; Pearson et al. 2005). Furthermore, differences between study teams and nonstudy teams in terms of performance improvement and other characteristics were generally modest and not statistically significant. It is likely, however, that study teams were more engaged in the collaborative and its features than nonstudy teams, and therefore not representative of nonstudy teams. Nevertheless, our study teams are representative of our population of interest: users of collaborative features. Study teams' familiarity with features makes them rich sources of information for a study assessing the helpfulness of features for those that use them. Other collaborative participants may have different views about the helpfulness of such features. For example, nonusers may view features as unhelpful for their situation. Thus, we caution against generalizing our findings to all collaborative participants. More work is needed to understand how nonusers view features.

Second, while our qualitative measure of performance improvement enabled us to combine data from collaboratives with different outcome measures, we lack objective, longitudinal measures of clinical practice, or outcome change. Third, although we found no evidence of evaluator bias in tests com-

paring collaborative-level improvement rates, the possibility of evaluator bias cannot be eliminated given our reliance on a single evaluator of performance per collaborative (i.e., the IHI Director). Finally, we only examined the BTS collaborative model. While BTS models are common, there are also several variations on this model (Solberg 2005). Additional studies are needed to evaluate those models in depth.

CONCLUSION

The objective of this study was to advance understanding about the collaborative model by examining participants' perspectives on its features. We found that participants value individual features differently, a helpful finding for collaborative sponsors, researchers, health care organizations, and policy makers interested in making collaboratives more effective for participants. Our results demonstrate that their effectiveness in facilitating significant performance improvement depends largely on their provision of features that foster interorganizational learning, in addition to features that foster intraorganizational learning.

ACKNOWLEDGMENTS

NOTE

1. We used 3.5 as our reference value because "the true limit of an interval (known as the real lower limit and the real upper limit) are the decimal values that fall halfway between the top of one interval and the bottom of the next" (Howell 1995, p.30);

3.5 fell halfway between 3, the score for moderate help, and 4, the score for great help. Therefore, 3.5 represented the real lower limit for great help and the real upper limit for moderate help.

REFERENCES

Ancona, D. G., and D. F. Caldwell. 1992. "Bridging the Boundary: External Activity and Performance in Organizational Teams." *Administrative Science Quarterly* 37 (4): 634–65.

Ayers, L. R., S. C. Beyea, M. M. Godfrey, D. C. Harper, E. C. Nelson, and P. B. Batalden. 2005. "Quality Improvement Learning Collaboratives." *Quality Management in Health Care* 14 (4): 234–47.

Berta, W. B., and R. Baker. 2004. "Factors that Impact the Transfer and Retention of Best Practices for Reducing Errors in Hospitals." *Health Care Management Review* 29 (2): 90–7.

Bradley, E. H., L. A. Curry, T. R. Webster, J. A. Mattera, S. A. Roumanis, M. J. Radford, R. L. McNamara, B. A. Barton, D. N. Berg, and H. M. Krumholz. 2006. "Achieving Rapid Door-to-Balloon Times: How Top Hospitals Improve Complex Clinical Systems." *Circulation* 113 (8): 1079–85.

Cohen, J. 1977. *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press.

Dellinger, E. P., S. M. Haussman, D. W. Bratzler, R. M. Johnson, D. M. Daniel, K. M. Bunt, G. A. Baumgardner, and J. R. Sugarman. 2005. "Hospitals Collaborate to Decrease Surgical Site Infections." *American Journal of Surgery* 190 (1): 9–15.

Fremont, A. M., G. Joyce, H. D. Anaya, C. C. Bowman, J. P. Halloran, S. W. Chang, S. A. Bozzette, and S. M. Asch. 2006. "An HIV Collaborative in the VHA: Do Advanced HIT and One-Day Sessions Change the Collaborative Experience?" *Joint Commission Journal of Quality and Patient Safety* 32 (6): 324–36.

Gilbert, D. T., and P. S. Malone. 1995. "The Correspondence Bias." *Psychological Bulletin* 117 (1): 21–38.

Glick, W. H. 1985. "Conceptualizing and Measuring Organizational and Psychological Climate: Pitfalls in Multilevel Research." *Academy of Management Review* 10 (3): 601–16.

Homer, C. J., P. Forbes, L. Horvitz, L. E. Peterson, D. Wypij, and P. Heinrich. 2005. "Impact of Quality Improvement Program on Care and Outcomes for Children with Asthma." *Archives of Pediatric and Adolescent Medicine* 159 (5): 464–9.

Horbar, J. D., J. Rogowski, P. E. Plsek, P. Delmore, W. H. Edwards, J. Hocker, A. Kantak, and P. Lewallen. 2001. "Collaborative Quality Improvement for Neonatal Intensive Care." *Pediatrics* 107 (1): 14–22.

Howard, D. H., L. A. Siminoff, V. McBride, and M. Lin. 2007. "Does Quality Improvement Work? Evaluation of the Organ Donation Breakthrough Collaborative." *Health Services Research* 42 (6): 2160–73.

Howell, D. C. 1995. *Fundamental Statistics for the Behavioral Sciences.* Belmont, CA: Duxbury Press.

Institute for Healthcare Improvement. 2003. *The Breakthrough Series: IHI's Collaborative Model for Achieving Breakthrough Improvement.* Boston: Institute for Healthcare Improvement.

James, L. R., R. G. Demaree, and G. Wolf. 1984. "Estimating Within-Group Interrater Reliability with and without Response Bias." *Journal of Applied Psychology* 69 (1): 85–98.

Janesick, V. 2003. "The Choreography of Qualitative Research: Minuets, Improvisations and Crystallization." In *Strategies of Qualitative Inquiry*, edited by N. Denzin and Y. S. Lincoln, pp. 46–79. Thousand Oaks, CA: Sage.

Jones, E. E., and R. E. Nisbett. 1971. *The Actor and the Observer: Divergent Perceptions of the Causes of Behavior.* New York: General Learning Press.

Kilo, C. M. 1998. "A Framework for Collaborative Improvement: Lessons from the Institute for Healthcare Improvement's Breakthrough Series." *Quality Management in Health Care* 6 (4): 1–13.

———. 1999. "Improving Care through Collaboration." *Pediatrics* 103 (1): 384–93.

Landon, B. E., L. S. Hicks, A. J. O'Malley, T. A. Lieu, T. Keegan, B. McNeil, and E. Guadagnoli. 2007. "Improving the Management of Chronic Disease at Community Health Centers." *New England Journal of Medicine* 356 (9): 921–34.

Landon, B. E., I. B. Wilson, K. McInnes, M. B. Lamdrum, L. Hirschhorn, P. V. Marsden, D. Gustafson, and P. D. Cleary. 2004. "Effects of a Quality Improvement Collaborative on the Outcome of Care of Patients with HIV Infection: The EQHIV Study." *Annals of Internal Medicine* 140 (11): 887–96.

Leape, L. L., G. Rogers, D. Hanna, P. Griswold, F. Federico, C. A. Fenn, D. W. Bates, L. Kirle, and B. R. Clarridge. 2006. "Developing and Implementing New Safe Practices: Voluntary Adoption through Statewide Collaboratives." *Quality and Safety in Health Care* 15: 289–95.

Marsden, P. V., B. E. Landon, I. B. Wilson, K. McInnes, L. Hirschhorn, L. Ding, and P. D. Cleary. 2006. "The Reliability of Survey Assessments of Characteristics of Medical Clinics." *Health Services Research* 41 (1): 265–83.

Mills, P. D., and W. B. Weeks. 2004. "Characteristics of Successful Quality Improvement Teams: Lessons from Five Collaborative Projects in the VHA." *Joint Commission Journal on Quality Improvement* 30 (3): 152–62.

Morse, J. M. 1994. "Designing Funded Qualitative Research." In *Designing Funded Qualitative Research*, edited by N. Denzin and Y. S. Lincoln, pp. 220–35. Thousand Oaks, CA: Sage Publications.

Morse, J. M., and L. Richards. 2002. *Readme First for a User's Guide to Qualitative Methods.* Thousand Oaks, CA: Sage Publications.

Pawson, R., and N. Tilley. 1997. *Realistic Evaluation.* London: Sage.

Pearson, M. L., S. Wu, J. Schaefer, A. E. Bonomi, S. M. Shortell, P. Mendel, J. A. Marsteller, T. A. Louis, M. Rosen, and E. Keeler. 2005. "Assessing the Implementation of the Chronic Care Model in Quality Improvement Collaboratives." *Health Services Research* 40 (4): 978–96.

Shortell, S. M., J. A. Marsteller, M. Lin, M. L. Pearson, S. Wu, P. Mendel, S. Cretin, and M. Rosen. 2004. "The Role of Perceived Team Effectiveness in Improving Chronic Illness Care." *Medical Care* 42 (11): 1040–8.

Simonin, B. L. 1997. "The Importance of Collaborative Know-How: An Empirical Test of the Learning Organization." *Academy of Management Journal* 40 (5): 1150–74.

Solberg, L. I. 2005. "If You've Seen One Quality Improvement Collaborative . . .." *Annals of Family Medicine* 3 (3): 198–9.

Strauss, A. L., and J. Corbin. 1998. *Basics of Qualitative Research: Techniques, and Procedures for Developing Grounded Theory.* Thousand Oaks, CA: Sage Publications.

Tucker, A. L., I. M. Nembhard, and A. C. Edmondson. 2007. "Implementing New Practices: An Empirical Study of Organizational Learning in Hospital Intensive Care Units." *Management Science* 53 (6): 894–907.

Van der Vegt, G. S., and J. S. Bunderson. 2005. "Learning and Performance in Multidisciplinary Teams: The Importance of Collective Team Identification." *Academy of Management Journal* 48 (3): 532–47.

Wilson, T., D. M. Berwick, and P. D. Cleary. 2003. "What Do Collaborative Improvement Projects Do? Experience from Seven Countries." *Joint Commission Journal of Quality and Safety* 29 (2): 85–93.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.