



**Center for
Effective
Organizations**

**Large Group Interventions: An
Empirical Field Study of their
Composition, Process, and Outcomes**

**CEO Publication
G 12-06 (613)**

Christopher G. Worley
*Senior Research Scientist
Center for Effective Organizations
University of Southern California
Marshall School of Business*

Susan A. Mohrman
*Senior Research Scientist
Center for Effective Organizations
Marshall School of Business
University of Southern California*

Jennifer Nevitt
*Director, Organization Development
Microsoft Corporation
Learning Group Consultant / Adjunct Faculty
Graziadio School of Business and
Management
Pepperdine University*

March 2012

**Large Group Interventions:
An Empirical Field Study of their Composition, Process, and Outcomes**

Christopher G. Worley, Ph.D.
Susan A. Mohrman, Ph.D.
Jennifer Nevitt, PsyD.

Center for Effective Organizations
Marshall School of Business
University of Southern California
Los Angeles, CA 90089
(213) 740-9814
<http://ceo-marshall.usc.edu>

With the generous support from Joel Fadem, Stu Winby, Kathryn Gaulke,
Alice Mark, Beth Neilsen, and Catherine Zhang

August, 2010

Large-Group Interventions: An Empirical Field Study of Their Composition, Process, and Outcomes

Abstract

Large group interventions are an important method of organization change. The large-group intervention literature is largely descriptive and normative, and contains a number of case studies that describe the process and some immediate outcomes. There is a large void with respect to empirical investigation. This research tested fundamental hypotheses related to large-group composition, process, and outcomes in a field study. Six large-group interventions (decision accelerators) were used to develop clinical service area strategies and instigate strategic change in a health care system. The results support the assertion that stakeholder diversity in the group's composition affects the number of stakeholder perspectives that were heard during the meeting and the breadth of issues addressed during decision-making, but failed to support the assertion that composition affects the intensity of debate and disagreement. Stakeholder diversity had a weak relationship with novel and relevant large-group outcomes, but debate intensity was strongly related to those outcomes. The implications of these results on large group intervention research and practice are discussed.

Large-Group Interventions: An Empirical Field Study of Their Composition, Process, and Outcomes

Some things don't change. Hamel (2007) noted that despite increasingly complex and rapidly changing environments, today's management processes and principles bear a striking resemblance to those of 50 years ago. But some things do change. In greater and greater numbers, organizations are implementing more agile, flexible, and changeable designs (Beer, 2009; Lawler & Worley, 2006; Pettigrew & Fenton, 2000), accelerating the learning and changing process (Tenkasi, Mohrman & Mohrman, 1998), and managing change more effectively (Beer, Eisenstat, & Spector, 1990; Kotter, 1996).

Among the practices addressing organization agility and innovation, large-group interventions (LGIs) are one of the fastest growing and most popular organization development interventions (Purser & Griffin, 2008; Cummings & Worley, 2009). Reflecting the complexity of the emerging environment, these interventions are built on principles of socio-technical systems (Cummings & Srivastva, 1977), self-organization (Wheatley, 1992), and high involvement and participation (Lawler, 1986). Despite their increasing popularity and the large volume of descriptive and normative literature (Bunker & Alban, 1997; Owen, 1992; Weisbord, 1992; and Dannemiller & Jacobs, 1992), there is a dearth of empirical research on large group processes and outcomes.

The purpose of this research was to test some of the primary relationships among large-group composition, process and outcomes proposed by LGI researchers and practitioners. As part of a larger change evaluation effort, the authors were able to study six LGIs set up to generate innovative clinical strategies for a health care system. The LGI's – a specific form of LGI known as a “decision accelerator” or DA – had a common format, common facilitators, and a common institutional context. Naturally occurring variation in their composition and in the processes that unfolded enabled us to look for systematic patterns among group composition, process, and outcomes of the LGI's.

Literature and Hypotheses

Large group interventions are one of the fastest growing organization development interventions (Bunker & Alban, 2005; Worley & Feyerherm, 2003; Cummings & Worley, 2009). LGIs are a

planned meeting or conference of organization members and other stakeholders to address organizational problems and opportunities. They can vary on several dimensions, including purpose (e.g., visioning, strategy formulation, organization design, or implementation planning); size (from less than 50 to over 2000 people); composition (mix of internal and external stakeholders); length (usually lasting between 2 to 4 days); structure (the agenda can be very rigid and formal or loose and informal); and number (single events or a series of linked conferences) (Holman, Devane and Cady, 2007). LGIs tend to have a common set of features, including a focus on the future, broad participation, an open-systems perspective, an intense period of planning prior to the meeting, information sharing, and a combination of divergent/creative as well as convergent/agreement processes (Bramson & Buss, 2002; Weisbord, 1987; Owen, 1992; Bunker & Alban, 2005; Manning & Binzagr, 1996).

The LGI literature is diverse and growing, but uniformly qualitative and case based (Garcia, 2007; Shmulyian et al., 2010)). To date, most research efforts have entailed (a) describing the large-group intervention methodologies defined by different practitioners or (b) describing individual case studies of specific large-group interventions. There is a strong bias toward coming up with the “latest” or “best” version as demonstrated by the numerous variations of large-group interventions, including Future Search, Conference Models, Simulations, Open Space, Work Out, Appreciative Inquiry Summit, Decision Accelerators, and World Café. As a result, the research base is full of thick descriptions of how to conduct a specific type of conference, what happened during a particular conference, and qualitative descriptions of the conference outcomes (e.g., “lots of energy,” “shared goals,” and “innovative strategies”) without much supporting evidence. Taken together, the case evidence suggests that LGIs can shorten decision-making cycle times, generate creative responses and strategies, and increase commitment to action (Cummings & Worley, 2009; Shmulyian et al., 2010).

Despite these strong assertions, and because each case may differ along one of several key dimensions, such as length, issue, method, facilitation, or agenda, there is very little empirical data, if any, on whether the basic premises of large group composition and process affect large group outcomes (Purser & Griffin, 2008). The research on LGIs has not verified important assertions about how their composition affects the internal processes or how those processes translate into outcomes. Such questions as “does the composition of the large group affect the level of debate?” or “under what conditions do large group interventions lead to

organization change?” have not been systematically examined. Interestingly, and in keeping with a practice bias among OD researchers, the literature on large-group interventions is more likely to compare the characteristics, processes, and assumptions of the different interventions than the outcomes produced (Bunker & Alban, 1992a; Shmulyian et al., 2010).

The intervention descriptions and case histories, however, do provide rough guidelines for testable hypotheses. For example, LGIs are rooted in open systems theory and assert that organization effectiveness derives from the explicit and tacit interactions among suppliers, customers, regulators, managers, employees, and other stakeholders. To change the effectiveness of a system therefore requires getting “the whole system in the room” (Weisbord, 1987; Axelrod, 1992). “Getting the whole system in the room” or bringing together diverse stakeholders is the most referenced design principle of LGIs. That is, as many different, relevant stakeholder groups as possible should be brought into the conference with the intent of bringing the maximum number of perspectives to bear on the issue, problem, or opportunity.

For any given purpose or agenda, having a group composed of diverse stakeholders is expected to lead to specific large-group processes (Weisbord, 1987; Bunker & Alban, 1997). The nature of the process will be sensitive to the composition. For example, if the group’s composition over-represents managers and executives, the conference is likely to represent the same kinds of conversation the organization always has. Similarly, if too many external stakeholders characterize the large group, there is little difference between an LGI and a focus group. The conference outcomes will likely reflect customer wants rather than a synergistic exploration of strategic and organizational alternatives.

However, the impact of group composition on large group process has not been empirically tested. Just because there are multiple stakeholder groups present in the room does not mean they will all interact, make inputs, surface their viewpoints and deal with conflicting preferences and perceptions, or consider an equally broad range of issues during decision-making. Practitioners know, for example, that unless the small groups that form and reform as part of the LGI are deliberately designed to represent the maximum mix of stakeholder groups (max mix principle), people from similar backgrounds will sit together and confer with each other. Similarly, Bunker and Alban (1997) specifically discuss the problem of “voice” because in large group settings, even when there are small group break out meetings, individuals may not believe they have had a chance to be heard.

In line with the assertions of large-group researchers and practitioners it is important to know if greater stakeholder diversity in the large group's composition is associated with greater participation or "voice" from the different stakeholder sectors that are represented, the breadth of issues considered in decision making, and the intensity of debate. Debate, in particular, is important to address since many LGI processes stress the formation of "common ground." The first set of hypotheses reflects the underlying assumptions about the impact of group composition on group process that are prevalent in the LGI literature:

H1a: The greater the diversity of stakeholders in an LGI, the greater the number of stakeholder voices that participate

H1b: The greater the diversity of stakeholders in an LGI, the broader the range of issues that will be considered in decision making

H1c: The greater the diversity of stakeholders in an LGI, the more intense the debate in conversations and dialogue.

The literature also supports the proposition that greater stakeholder diversity and broader and more intense interactions lead to superior large-group outcomes. Organizations adopt LGI processes because they promise to generate creative and innovative outputs, accelerate change processes, and build change capabilities (Bryson & Anderson, 2000; Bunker & Alban, 1992b, 1997; Weisbord, 1987). The large-group literature's concern with involvement from diverse stakeholders is to increase the probability of novel and relevant outcomes (Weisbord, 1987; Bunker & Alban, 1997). The more diverse the stakeholder composition of a large group, the more the outcomes should be novel and relevant. Our second set of hypotheses predicts a direct impact of composition on outcomes:

H2a: The greater the diversity of stakeholders in an LGI, the more novel the outcomes.

H2b: The greater the diversity of stakeholders in an LGI, the more relevant the outcomes.

LGI scholars and practitioners also posit that the greater the number of stakeholder voices that are heard and the broader the range of issues discussed, the more novel and relevant the outcomes should be (Weisbord, 1987; Bunker & Alban, 1997). For example, in their case study of an IKEA redesign, Weisbord and Janoff (2005) note, "Some 52 stakeholders examined the existing system, developed a new system, created a strategic plan, and formed task forces led by

key executives to implement it. In 18 hours, the plan was made, validated, and signed off by the company president, key people from all affected functions, and several customers” (p. 71).

Because the LGI raises more issues, positions, and points of view; recognizes the unique strengths of the firm or unique opportunities in the environment; and reveals more of the required urgency for action, the LGI is more likely to generate novel and relevant outcomes. The third set of hypotheses deals with the impact of these process variables on outcomes:

H3a: The greater the number of stakeholder voices that participate, the more novel the outcomes

H3b: The greater the number of stakeholder voices that participate, the more relevant the outcome

H4a: The broader the range of issues considered in conversations, the more novel the outcomes

H4b: The broader the range of issues considered in conversations, the more relevant the outcomes.

Finally, the large-group literature further asserts that intense, deep, and meaningful discussions should reveal nuances, produce insights, and generate knowledge that are translated into more novel and relevant outcomes (Danemiller & Jacobs, 1992; Miller, Fitzgerald, Murrell, Preston, & Ambekar, 2005; Barrett, Thomas, & Hocevar, 1995). By sharing power, information, knowledge, and skill (Lawler, 1986), large groups foment conditions for innovation. For example, Barrett, Thomas, and Hocevar (1995) found that deep discourse produced innovative strategies and successful organization change. Miller et al. (2005) found that relatively shallow conversations and the lack of conflict contributed to a narrow range of issues being discussed, less creative strategies, and a sluggish implementation process in a joint venture design process. Similarly, Tenkasi, Mohrman, & Mohrman (1998) found that dialogue characterized by both positive and negative conversations led to the development of shared understandings and fostered the learning necessary for a system to implement fundamental change. Therefore, we propose two more hypotheses relating LGI process to outcomes:

H5a: The more intense the debate in an LGI, the more novel its outcomes.

H5b: The more intense the debate in an LGI, the more relevant its outcomes.

Methods

Research Setting

Alegent Health is a not-for-profit, faith-based healthcare system and the largest private employer in Nebraska. It was established in January, 1996 through a joint operating agreement between Catholic Health Initiatives and Immanuel Health Systems. Headquartered in Omaha with a service area extending to southwestern Iowa, Alegent Health in 2007 operated five large, acute care hospitals, four regional hospitals, and more than 100 sites of service with over 1,300 physicians on its medical staff and approximately 9,000 employees. Its founding mission and vision statement, crafted in 2004, relied on its faith-based heritage and challenged the organization to achieve a world-class standard.

In 2004, a new CEO was appointed to lead Alegent Health with several clear mandates from the board. The most pressing issue was the repair of the distressed relationships between the organization and its attending physicians. With the exception of some of the primary care physicians in the outlying medical clinics, the physicians were not employees of Alegent, but rather community physicians with hospital privileges. The prior administration had alienated the physician community, and a public vote of “no confidence” that was played out in the local media had damaged Alegent’s reputation.

In addition to emergency medicine and primary care, the organization operated six “service lines.” The women's and children's services, cardiovascular services, orthopedics, oncology, neurosciences, and behavioral health services had been the drivers of strategy and operations throughout the system. As part of the CEO’s overall change agenda, he reorganized Alegent giving the hospital presidents more operational control of the facilities. The service line vice presidents were charged with responsibility for setting strategy and working with the presidents to execute it.

The CEO also created the position of Chief Innovation Officer, a new position charged with invigorating the service lines. A former consultant who had worked on the reorganization filled the CIO position. As the CIO considered how to advance innovation in the organization, he brought on two organization consultants who introduced him to large group interventions and decision accelerators (DAs). After several site visits to organizations with reputations for innovation and the use of large-group interventions, Alegent committed to this technology.

Dubbed the “Right Track” (RT) process, its purpose was to spark strategic and organizational innovations in the system and support the organization’s pursuit of world-class healthcare.

A DA is a type of large-group intervention (Shmulyian et al., 2010). Like other LGIs, the DA brings multiple stakeholders together over multiple days to address an organizational issue, and uses a divergent and convergent process of decision-making. The DA differs slightly from other types of LGIs in that it focuses on a process of “rapid prototyping.” This process involves small “max mix” groups working on a specific task, such as creating a vision or strategy, which is then presented to the large group and that “product” is then iterated based on different perspectives (e.g., stakeholders) and inputs (e.g., business and market expectations).

The CIO organized six initial DAs in July or August of 2005, one for each of the service lines. The purpose of each DA was to generate an innovative vision and core strategies for that service line. While customized to some extent by the service line VP sponsor, the agenda, process, and activities of each DA were remarkably similar. Each included an identical environmental scanning activity, process for discussing the definition and characteristics of being a “world class” organization and service line, activities related to the creation of a service line vision, and an exercise designed to identify high level milestones of vision accomplishment over three time horizons (Baghai, Coley, & White, 2000). Table 1 provides a sample agenda and the extent to which each DA utilized a particular exercise or activity. One of the two organization consultants always served as the lead facilitator of the DA.

Table 1
DA Agenda Comparison

DA Activity or Exercise*	Oncology	Orthopedics	Neuroscience	W&C	Cardiology	Behavior
Mapping the forces affecting	X	X	X	X	X	X
Implications and scenarios of the forces	X	X	X	X	X	X
Exploring the implications of world class to Alegent	X	X	X	X	X	X
Exploring the implications of a world class vision on the service line	X	X	X	X	X	X
Ranking service line vision elements	X	X	X	X	X	X
Developing a technology timeline		X	X	X	X	

Key issues exercise	X	X	X	X	X	X
Conference review	X	X	X	X	X	X
Horizons and milestones activity	X	X	X	X	X	X
Summary presentations to ELC members	X	X	X	X	X	
Unique activities						Leverage the patient exercise

*For any particular exercise or activity that required both large and small group tasks, the small groups were always set up according to “max-mix” principle that assigned as many different stakeholder groups as possible.

In support of initiatives springing from the original six DAs and other strategic issues in the system, Alegent Health conducted over 100 DAs between July, 2005 and March, 2007, and continues to use them in support of strategic change. In June, 2008, Alegent was recognized by the Network for Regional Healthcare Improvement (NRHI) as the number one health care organization in the nation in terms of clinical/quality outcomes and patient satisfaction.

The Research Process and Sampling

As Alegent Health approached the 18-month mark following the original six DAs, a group of university-based researchers (not the consultants) proposed to assess the impacts of the DA on the organization and provide recommendations about how to build capabilities for the future. The opportunity to study the dynamics of composition, process, and outcomes presented itself during this evaluation process. With the facilitators, format, and organizational context constant across all six DA’s, it was possible to take advantage of the naturally occurring variation in their composition and process to investigate the impact of composition and process on the outcomes of the LGI’s.

The overall assessment was conducted in three phases, and the data for this study were collected during the assessment. The *first phase* consisted of interviewing Alegent Health

executives and physicians, and reviewing archival data to learn about the system, their expected outcomes, and the DA interventions.

The *second phase* involved the collection and analysis of two data sets. The first data set consisted of transcripts and reports from the six service line strategy DAs. For each DA, the Right Track staff assembled all participant names and affiliations, agenda activities, elapsed times and instructions for each activity, and photographs of different activities and all small group outputs. In addition, two to three Right Track staff members sat in the back of the meeting room and transcribed all large-group report out presentations and ensuing discussions and debates. To the extent possible by a group of experienced computer users, the transcripts represent a verbatim record of these discussions. These DA reports yielded data related to group composition and group process as described below. The report did not include comments, observations, or transcripts of small breakout-group meetings, so that aspect of the DA process was only available through their report-outs to the large group.

The second data set consisted of structured interviews and survey responses from 31 executives, hospital presidents, service line VPs, nursing directors, and physicians in December, 2006 (Appendix A). All but three of these 31 individuals participated in more than one DA, and were responsible for working in the service line, developing and executing the service line strategy, or managing functions that support the Alegen system. Thus, each individual was an appropriate and legitimate source for rating the quality of the outcomes produced by the DA, and because many of them attended more than one DA, they were also a good source of comparison across events.

Twenty-six of the interviewees were executives and represented 100% of the senior managers in the organization and 60% of the hospital presidents. Two to three physicians from each specialty were invited to participate. Of the 11 interviews that were scheduled, only five physicians completed interviews primarily due to emergencies but also because of oversight and schedule conflicts (double booking). Two of the physicians had attended two DAs.

In service of the overall assessment, the interviews focused on the broad strategic changes taking place at Alegen Health. An interview protocol of 15 questions was developed that asked about Alegen's mission and strategy, recent organization changes, purpose of the DA process, characteristics of particular DAs, and the overall progress of implementation. As part of

the interview process, the participants were asked to complete a short survey about the characteristics of the service line strategy for any DA they attended or were familiar with.

The *final phase* of the study involved coding, analyzing, and integrating the data sets and feeding back the information to a group of Alegant Health executives, including the Chief Innovation Officer, service line VPs, Right Track program office staff, and other senior managers. This meeting and discussion were important in understanding and interpreting the patterns of data, and these interpretations are included in the results section.

The purpose of the present study – to test the relationships among composition, process, and outcomes of a large-group intervention – supplemented the purpose of the overall assessment. The multiple sources of data that became available (transcripts, interviews, and products), coupled with our knowledge of the literature and interest in LGI approaches, presented a unique opportunity to address a gap in the LGI research. The primary data for this study comes from the survey completed by executives and physicians and the transcripts of the individual DAs. Where possible, data from the interviews were used to corroborate, triangulate, and improve reliability.

Measurement

To address the hypotheses, three categories of variables were operationalized: large group composition, large group process, and large group outcomes. Each of these is described in turn.

Large group Composition Variables. According to the literature, the key composition issue for an LGI is stakeholder diversity or “getting the whole system in the room.” The DA transcripts contained the names, titles, and affiliations of each participant, and we developed a direct measure of the DA’s composition based on this list. The relevant range of stakeholders included the senior Alegant managers (e.g., hospital presidents, clinical area VPs, and other operating executives) responsible for allocating resources and setting initiatives for strategy execution and the physicians who would be asked to deliver clinical services in a certain way and with a certain standard of quality. Other stakeholders with an interest in a clinical strategy were patients, regulators, community members, and government representatives. Thus, we divided the range of potential stakeholders into three groups – Alegant managers, physicians, and the larger community interests. For this classification, the small group of physicians employed by Alegant as primary care doctors was included in the Alegant manager group.

The *Stakeholder Diversity* variable measured the extent to which the three stakeholder categories (Alegent managers, physicians, and community members) were equally represented. The more equally represented, the more diverse the group. A “3” was assigned if each stakeholder category was equally represented. For example, the Behavioral DA had 39% employees, 27% physicians, and 34% community members. A “1” was assigned if one or two stakeholder categories were over or under represented. For example, the Orthopedics DA had 48% employees, 44% physicians, but only 8% community members. All other DAs were coded a “2.” This direct measure also afforded us the opportunity to calculate the percentage of DA attendees who represented the two external stakeholder groups.

The *Percent Physicians* variable measured the proportion of total attendees that were on-staff physicians but not employees of Alegent Health. In light of the recent history of physician mistrust of Alegent and its goal to improve relationships with physicians, these were important stakeholders in the process and their effect on LGIs was worth exploring. Moreover, measuring this aspect of group composition was deemed important given the power of physicians and their criticality in effecting change in healthcare in general.

The *Percent Community members* variable measured the percentage of participants who were community members, including patients, government representatives, customers, vendors, non-profit organizations, or health care regulators. This aspect of group composition was deemed important given the importance of the open system perspective to large group interventions (Bunker & Alban, 1997) and the role that non-central stakeholders might play. (Note: By definition, the percent of Alegent members in the DA equals 100 minus the sum of the above two variables.)

Large group Process Variables. Three process variables were defined and coded from the transcripts. Two coders – one of the authors and a graduate of a master’s degree program in organization development – met to discuss the purpose of the study and the coding guidelines. The coders independently rated and then discussed a randomly selected DA report to assure common understanding of the coding categories. The other five DA reports were coded independently and compared later for consistency. Differences in coding were discussed and resolved using the definitions as a guide and attempting to remain consistent across the cases. Reliability statistics were computed prior to the resolution of differences to provide a baseline measure of consistency (presented below).

The *Number of Stakeholder Voices* variable represented the number of different stakeholder categories who participated in debriefs and discussions. For each activity report out or large-group discussion, we identified the number of different stakeholder groups whose representatives spoke. The discussion was rated on a 1 to 5 scale where a “1” represented little or no participation from any stakeholder group beyond the formal report outs and a “5” represented more than three different stakeholder groups involved in the discussion. [Note: Even though the stakeholders were grouped into three categories for the stakeholder diversity variable, the detail in the transcripts allowed us to be quite specific about who was talking – physicians, executives, patients, government officials, etc.] The number of stakeholder voices value for each DA was the average rating across the different discussions, and represented an overall indication of how many different stakeholder voices were heard during the DA.

The *Range of Issues Considered* variable reflected the breadth of topics considered (no matter who raised them) during discussion and decision-making. It measured the average number of issue-oriented categories mentioned and considered during three activities where the service line vision was iterated. The final list of categories represented a blend of both theoretical and practical considerations (a list of the categories is available upon request). Theoretically, several different categories were derived from the strategy literature, including dependence on technology, marketing requirements, and the need for partnerships. Practically, the goals of the Alegent Health strategy also drove the categories, including such categories as the link to the mission/vision of Alegent and the extent to which the hospitals and other Alegent facilities were coordinated and integrated.

The *Debate Intensity* variable captured a qualitative aspect of the DA’s process, specifically the extent to which disagreements were passionate, deep, or intense. To achieve this measurement, we looked at the amount of words spoken, the emotional intensity implied by different words (e.g., did the participant say they had a “different view of things” or did they say they were “upset” with the issues being discussed?), and the extent to which perspectives expressed by one person were agreed to or disputed by another. All of these indicators contributed to an assessment of the DA’s debate intensity.

Researchers coded incidents of debate or conflicting viewpoints expressed in the large group report outs and discussions. For each activity report out/large-group discussion that occurred during each DA, the intensity of the debate was rated on a 1 to 5 scale. A “1”

represented a low level of debate, where, for example, a presenter was asked a question for clarification but there was little discussion. A “3” represented moderate debate where there were not only questions but the expression of mild disagreement or the offering of an “alternative” point of view. A “5” represented a high level of conflict where the words indicated considerable emotional charge, the discussion often exceeded the time allowed for the exercise, and “fundamental differences” were expressed and debated. The debate intensity value for each DA was the average rating across the different discussions, and represented an overall indication of the extent of disagreement expressed and discussed during the DA.

Coding for emotional intensity from a written transcript has certain inherent hazards. In an attempt to provide some validity to the ratings, we consulted answers to the interview question, “For any DA you attended, what were the biggest tensions in the process?” Five interviewees – the two consultants and three executives – attended four or more of the DAs and thus had a basis of comparison across DAs. These five interviews form the database for this triangulation. Each interview response was reviewed for words and phrases indicating the amount or type of tension present and then a relative ranking of high, medium, or low was assigned for each DA attended. This provided a different perspective on the debate intensity in the DAs.

There was very good correspondence between the coded ratings and the interview assessments. The interviewees ranked the two DAs that had been coded as having the lowest debate intensity lowest in tension; the interviewees ranked the two most conflictual DAs high and medium; and they ranked two moderately rated DAs high and medium. In combination with the comments provided by managers during the feedback session (described below), there is good reason to believe that this measure of debate intensity conforms well to the experiences of the organization.

The baseline inter-rater agreement for the number of stakeholder voices and debate intensity variables was 0.75, an acceptable level of reliability. Where raters disagreed by more than a point on the scales, a discussion took place to agree on a rating. Where the disagreement was only one point, the average of the two ratings was used as the score. For the range of issues considered variable, the baseline inter-rater agreement across all DAs and ratings was 0.93.

Outcome Variables. The primary output of the DA was a service line vision and strategy. To address the novelty and relevance of the outcomes, the survey measured one novelty scale

and three relevance scales. Novelty represents the distinctiveness and innovativeness of the strategies (Fagerber, 2004). Comprehensiveness (Fredrickson and Mitchell, 1984), aggressiveness (Brittain and Freeman, 1980), and congruence (Ledford, 1984) were expected to represent the relevance dimension. Comprehensiveness was chosen because the organization was utilizing the DA process to account for a broader range of issues in its formulation process, including especially the concerns and perspectives of the physician community. Aggressiveness was chosen to represent the urgency being felt in the system. Congruence was chosen because the organization expected the strategies to be implemented and they were interested in the extent to which the strategies were congruent with other aspects of the health care system’s activities.

Table 2 presents the items used to measure the strategy dimensions, the correlations among the items, and each scale’s Cronbach reliability estimate. When these eight items were submitted to a confirmatory factor analysis, only one factor emerged suggesting more commonality among the items than intended. As a result, an overall “novel and relevant” strategy scale, composed of all eight items was formed. The four intended scales are retained to understand how the variables might affect the more nuanced aspects of strategy.

Table 2
Decision Accelerator Outcome Variables’ Intercorrelations and Scale Reliabilities

		Reliability	1	2	3	4	5	6	7	8
	Comprehensiveness	0.91								
1	It accounts for the key issues facing this service line		1.00	0.83	0.69	0.68	0.68	0.70	0.69	0.64
2	It addresses an appropriate range of market concerns			1.00	0.66	0.61	0.64	0.68	0.65	0.66
	Aggressiveness	0.80								
3	It implies an appropriate urgency				1.00	0.67	0.68	0.65	0.61	0.61
4	The goals are bold enough					1.00	0.73	0.63	0.68	0.49
	Novelty	0.86								
5	It differentiates us from the competition						1.00	0.76	0.68	0.62
6	It leverages our unique strengths							1.00	0.73	0.73
	Congruence	0.83								
7	Supportive of the health system's purpose and culture								1.00	0.70
8	Aligned with other changes taking place in the system									1.00
	Novel and Relevant Outcomes (all 8 items)	0.94								

N = 73, all correlations significant at the p<.05 level

The restructured organization was already beginning to manifest several positive business outcomes, such as improved physician and patient satisfaction, market share, and community relations. Because the explicit purpose of the DAs was to generate novel and relevant strategies to achieve world-class health care delivery, not to plan implementation or drive business results, our focus in this article is on the extent which DA composition and process affected this intended DA outcome.

Analysis

The results from the DA report coding were entered into a spreadsheet, along with aggregated service line data from interviews and surveys. Acknowledging the small sample size, two analytic concessions were made. First, we utilized non-parametric, Spearman rank order correlation coefficients to test the hypotheses. Second, a liberal significance level was established at $p = .15$. We reasoned that a typical Type I error rate of .05 would not only be difficult to achieve given the small sample size, but might risk missing existing patterns in the data. On the other hand, a too liberal p level might result in chasing indefensible relationships. These concessions seemed reasonable under the circumstances. This is the first empirical comparison of LGI dynamics and the likelihood of finding a larger number of comparable LGIs seems very low. Practically, the small degrees of freedom in the rank order correlation calculation means that even small differences in the rankings can quickly lower the result. We chose the high alpha level to facilitate the identification of patterns that might set a baseline for future research.

Results

Descriptive Statistics

As shown in Table 3, the DAs varied in their composition, process, and outcomes. For example, stakeholder diversity was greatest in Behavioral Health and least in Orthopedics. The percent of physicians in the DAs ranged from 47% in the Oncology DA to 27% in the Behavioral Health DA. The percent of community members ranged from 34% in Behavioral Health to 8% in Orthopedics.

The number of stakeholder voices was lowest in Orthopedics and highest in Behavioral Health. The average rating of 2.00 for Orthopedics suggests that, on average, transcribed discussions evidenced much discussion mostly among one group, while the 3.11 for Behavioral

Health suggests discussion involving two to three different stakeholder groups. Behavioral Health and Neuroscience considered the broadest range of issues in developing their vision with an average of 6.6 and 6.4 categories presented while Orthopedics considered a narrower set, averaging 4.1 categories.

Table 3
Characteristics, Composition, Process, and Outcomes of the Original DAs

	Oncology	Orthopedics	Neuroscience	W&C	Cardiology	Behavior
Composition						
Stakeholder diversity	2	1	2	2	2	3
Percent of attendees who were physicians	47%	44%	38%	39%	41%	27%
Percent of attendees from community	25%	8%	17%	16%	21%	34%
Process Characteristics						
Number of Stakeholder Voices	2.21	2.00	2.72	2.17	2.75	3.11
Range of Issues Considered	5.7	4.1	6.4	4.8	5.2	6.6
Debate intensity	1.14	0.69	0.56	1.17	1.40	1.44
Strategy Outcomes						
Relevance: Comprehensiveness	4.30	3.71	3.71	4.13	4.11	4.12
Relevance: Aggressiveness	3.90	3.04	3.21	3.94	3.79	4.00
Relevance: Congruence	3.77	3.08	3.21	3.94	3.57	3.77
Novelty	3.63	3.04	3.21	3.69	3.71	4.04
Novel and Relevant Outcomes	4.01	3.32	3.43	3.97	3.90	4.10
Number of DA Participants	57	48	42	57	56	44
Survey Sample Size	14	9	15	12	12	13
Percent of participants who rated outcomes	24%	19%	33%	21%	21%	29%

Debate intensity was relatively low overall, but was considerably lower in Neuroscience and Orthopedics and higher in Behavioral Health and Cardiology. At the feedback session, organization managers took issue with these results. They believed that the level of debate and conflict within any given DA was quite high. The managers were concerned that the scores reflected the coding of written *words and phrases* and may not have registered the affect visible in the live situation. However, there was good agreement that the *relative* debate intensity reflected the different service line DAs well. That is, everyone in the meeting agreed that the

Behavioral and Cardiology service line DAs were more contentious than Neuroscience and Orthopedics. This gives confidence in our use of rank order correlations while at the same time pointing out the limitations of coding emotive variables from transcripts.

Orthopedics and Neuroscience had lower ratings on the outcome dimensions and the overall novel and relevance scale. Women/Children's, Behavioral Health, and Oncology scored consistently higher. Given the timing of the surveys (December, 2005), the lower scores for Orthopedics and Neuroscience may have been influenced by actions taken by orthopedic physicians to begin their own specialty hospital and a leadership transition that occurred in the VP position that manages both of these service lines. While these actions are more likely to impact implementation, it's possible that they influenced perceptions of the strategy's dimensions as well.

Table 3 also indicates survey sample sizes for each DA and the percentage of DA participants the sample size represents.

Hypothesis Testing

The first group of hypotheses proposed that stakeholder diversity in a large group's composition would be positively related to the number of stakeholder voices heard, the range of issues considered, and debate intensity. As shown in Table 4, stakeholder diversity was positively related to all three process variables but significantly ($p < .10$) related to the number of stakeholder voices and the range of issues considered. There is good support for H1a and H1b, but no support for H1c.

The two other composition variables provide additional insight into this result. For example, higher percentages of physician attendees were negatively related to all process variables and significantly related ($p < .10$) to the number of stakeholder voices. On the other hand, the percent community members variable was positively correlated with all process variables ($p < .15$ level), including debate intensity.

Table 4
Relationships between DA Composition and DA Processes

	Number of Stakeholder Voices (H1a)	Range of Issues Considered (H1b)	Debate Intensity (H1c)
Stakeholder Diversity	0.82*	0.81*	0.65
Percent Physician Attendees	-0.80*	-0.58	-0.35
Percent Community Attendees	0.75 ^t	0.76 ^t	0.74 ^t

Significance of Spearman Rank Order Correlations: ^t < .15; * < .10; ** < .05; *** < .01
N=6

The second set of hypotheses proposed that stakeholder diversity would be related to novel and relevant outcomes. As shown in the first row of Table 5, the stakeholder diversity variable was positively and significantly related to novelty ($p < .05$). The overall novel and relevant scale was also related but only at the $p < .15$ level. The two other composition variables again provide insight. Four out of five of the Spearman correlations between percent physician attendees and the outcome variables were negative but none of them were significant. The percent of community representatives was positively associated with all strategy variables, and significantly related to more aggressiveness ($p < .15$), novelty ($p < .05$), and overall novelty and relevance ($p < .10$). The results support a relationship between stakeholder diversity and the novelty or distinctiveness of LGI outcomes (H2a).

Table 5
Relationships Among DA Composition, Process Diversity, and DA Outcomes

	Relevance: Comprehensiveness	Relevance: Aggressiveness	Relevance: Congruence	Novelty	Novel and Relevant Outcomes
Stakeholder Diversity (H2)	0.53	0.73	0.64	0.86**	0.76 ^t
Percent Physicians	0.03	-0.29	-0.22	-0.54	-0.30
Percent Community	0.65	0.76 ^t	0.61	0.88**	0.80*

Number of Stakeholder Voices (H3)	0.10	0.32	0.14	0.57	0.35
Range of Issues Considered (H4)	0.16	0.31	0.20	0.45	0.35
Debate Intensity (H5)	0.83**	0.90**	0.79*	0.93***	0.91**

Significance of Spearman Rank Order Correlations: $t < .15$; * $< .10$; ** $< .05$; *** $< .01$
N = 6

The third and fourth sets of hypotheses proposed that number of the stakeholder voices and the range of issues considered would be positively related to the novelty and relevance of outcomes. As shown in Table 5, both of these process variables were unrelated to the outcome variables. There is no support for H3 and H4.

Finally, H5a and H5b are supported. The final row of Table 5 shows that there are positive and significant relationships between debate intensity and all strategy scales.

Discussion

The purpose of this research was to empirically test relationships among large-group composition, process, and outcomes touted in the descriptive and case based LGI literature. The results of this study support some of those relationships but also suggest more complexity than expected.

The most common assertion of LGI proponents is that “getting the whole system in the room” (Weisbord, 1987) will lead to richer, more meaningful discussions, better strategies, faster change, and improved effectiveness. Hypotheses 1 and 2 were intended to address the first two of these claims. That is, does a large group’s composition affect processes and outcomes? Hypothesis 1 sought to verify a relationship between the diversity of stakeholders represented and large-group process characteristics. This hypothesis was confirmed when process was measured by the number of different stakeholder voices heard in discussions (H1a) and the range of issues considered during decision making (H1b) but not when process was measured by debate intensity (H1c). A closer look suggested an important twist on the “get the whole system in the room” principle.

“Getting the whole system in the room” is a catchy phrase that has multiple interpretations and may be difficult to operationalize. When stakeholders are broadly categorized

as they were here, getting the whole system in the room is more than simple representation; it must be balanced representation. Too high a concentration of a stakeholder can tilt the conversation. The number of stakeholder voices, the range of issues considered, and debate intensity were *negatively* related to the percent of physician attendees, although only one relationship (number of stakeholder voices) was significant. On the other hand, higher percentages of community attendees, which were generally 50% less than physician percentages, were *positively* related to debate intensity and range of issues considered. Together, these results suggest that too much of one stakeholder group was negatively associated with good large-group processes, and higher concentrations of under-represented stakeholders was positively associated with good process. The most direct measure of stakeholder representation (stakeholder diversity) supports the importance of having a balanced composition. When there was good balance in the representation of different stakeholder groups, there was good representation of these groups' perspectives in conference discussions of vision and strategy.

An additional explanation for that pattern of findings relates to the central and traditionally high power role of the physician in health care settings. Overrepresentation of the physicians compared to other groups may have contributed to a group process in which some voices were not heard. In the Alegent DAs, the high concentration of physicians was partly a function of one of the goals of the intervention – to address tensions in the relationships with physicians. Subsequent improvement in physician relations that were reported in the interviews may support the wisdom of this approach. Nevertheless, the results here clearly suggest that balanced representation of stakeholder perspectives results in more comprehensive and diverse points of view being raised in the DA. This result seems particularly robust given that the other participants in the DA had roughly equivalent power bases. The Alegent participants were hospital presidents, clinical area VPs, and other senior executives from the system; the community members were regulators, patients, and representatives from local government.

Stakeholder diversity did *not* relate to greater debate intensity. There was virtually no evidence that balanced representation resulted in more heated discussions. The only exception was a weakly significant relationship between community percentage and debate intensity. Just because there are diverse perspectives in the room and in the conversations does not mean that the discussion of those perspectives will be more or less intense and meaningful. We discuss this relationship further below.

In hypothesis 2, we found no support for the relationship between stakeholder diversity and outcome relevance (H2b) but good support between stakeholder diversity and novelty (H2a). This suggests that having good, balanced representation from all stakeholder groups facilitated the development of strategies that were differentiated from the competition and leveraged unique organizational strengths. As with H1, it appears that the presence of community members was most helpful in reaching this novelty. Developing novel and distinctive strategies may rely on the presence and influence of several points of view and balanced representation may be necessary for that to happen. In the case of health care, community members may have the voice least likely to be heard and the most likely to contribute to a unique strategy. Including them in strategizing increased the likelihood of novel strategies.

It is particularly curious that there was no relationship between stakeholder diversity and comprehensiveness. Given that stakeholder diversity was associated with more voices being heard and more issues considered in decision making, expecting a relationship between stakeholder diversity and comprehensiveness would seem reasonable. Looking at H3, H4, and H5 sheds some light on this issue.

H3, H4, and H5 tested whether large group processes affected large group outcomes. There was no support for H3 or H4. From a process perspective, just hearing a diversity of perspectives or considering a broad range of issues had no impact on outcomes. In particular, the range of voices heard and the breadth of issues considered were not related to any of the outcomes, including comprehensiveness. It is possible that too high a concentration of physicians narrowed participation and expression of alternative points of view. Only when debate intensity was high (see below) did relevance and novelty emerge.

However, there was strong support for H5. Debate intensity in the DA was an important driver of novel and relevant strategies. Whereas stakeholder diversity was related only to novelty, debate intensity was positively related to all of the strategy dimensions and appeared as the strongest driver of DA outcomes. When different preferences and perspectives were debated more, novel and relevant strategies were produced. When a broad set of issues were not only heard but considered, assessments of novelty were more likely to result from discussions that accounted for more diverse knowledge and experience, calibrated the appropriate amount of urgency and aggressiveness, and clarified the way the strategy aligns with the organization's mission/values and other changes. The LGI literature downplays the role and importance of

debate and conflict – one important source of creative and new approaches. The data here suggest that having open debate really helped.

The importance of debate intensity is not inconsistent with the finding regarding composition (stakeholder diversity). Conflicting and intensely held preferences and viewpoints may be more likely to find expression in a setting where the various stakeholders are balanced. If there are not enough numbers of each stakeholder, their voices are less likely to be heard. The suppression of views may be exacerbated if the under-represented group is traditionally less powerful in the system. But balanced representation alone may not lead to open disagreement, for example when all stakeholders have common understandings of the situation. On the other hand, lack of balance may well prevent some underrepresented stakeholders from expressing their views openly and with conviction.

Limitations

This study had several limitations. The most obvious limitation is the small number of large-group interventions analyzed. The small sample size in this study could bias the results in a variety of ways, including a restricted range of values for any particular variable, little confidence that variables are normally distributed, and non-random sampling challenges to representativeness and our ability to generalize the results to the larger population of LGIs. The small sample size clearly limited the statistical options that might have been able to illuminate a variety of moderating forces and influenced our decision to adopt a liberal significance level (although none of our conclusions depend on that concession).

However, the positive relationships found in this study warrant consideration; a variety of significant alternative explanations was held constant and can be ruled out. That is, the facilitators were the same, the agendas were identical, and the conferences were held in the same facility, within the same context, and in the same time frame. In addition, the diversity of data collection times, sources, and methods used in this study obviates concern about common-method issues. Our sample size is small, but the field study provided several excellent and naturally occurring controls. The likelihood of finding a field situation with a larger number of LGIs seems remote.

That said, additional limitations should be noted. There may be several explanations for why variety of voices, broad consideration, and conflict intensity were more or less related to

group composition and outcomes. All of these variables were coded from the DA transcripts, which represented a rich and detailed, but not perfect, data source. First, the group process variables were only based on large-group discussions and report outs. Important debates, conflicts, and issues may have taken place in the small group exercises that reduced the variety and intensity of discussion in the large group. Second, the person reporting out the results from a small group discussion could also affect the large group process. A well-respected administrator, for example, might be able to affect the variety of voices heard or suppress conflict. (We are grateful to the reviewers for raising this issue.)

Finally, the data used in this study were originally part of an overall evaluation of the DA's role on health care system outcomes and in creating an innovation capability. Our assignment and interest was to look at this broader issue and not at the specific intra-DA dynamics. As the evaluation unfolded, the advantages of a field study with naturally occurring controls became clear. Our awareness of the dearth of empirical research on the basic assertions of a popular OD intervention made the present study worth pursuing. The limitation is that we were somewhat stuck with the data we had. Although the DA transcripts were a robust data source, they had limitations noted above. Similarly the way we asked people to complete a survey for each DA they attended turned out to be an advantage, but only in retrospect. Thus, our judgment was that the benefits to be derived from "working with what was available" exceeded the limitations of the data.

If we had to do it all over again, we would have developed a better survey sample of the DA participants (in particular gathering the views of participants in the "community" category), and a more comprehensive set of questions that could be triangulated with the interviews. Similarly, we could have developed a more precise interview protocol with the intention of teasing out whether participants saw any connections between stakeholder diversity and the group processes and outcomes. We might then have been able to provide a richer qualitative description of what was happening in the DAs.

There are, then, clear steps that can be taken to improve the quality of the data collected with respect to the hypotheses of the study. If researchers were to take on the task of replicating or extending this study, there is an obvious way forward. However, that doesn't affect the core limitation and the core strength of this study. We still would have only had six interventions to

work with, and the likelihood of finding multiple large-group interventions with as many naturally occurring controls seems terribly small.

Suggestions for Future Research

Although we believe some important and basic relationships in the LGI research have been addressed here, there is still much to be learned. For example, we still need to know whether the energy and momentum generated by an LGI translates into actual organization change or improved performance. The subsequent changes at Alegen Health and their improved performance provides one positive data point in that arena.

With respect to the internal dynamics of LGIs, the results of this research leave at least one important question unanswered. Given the role of debate intensity on the strategy outcomes and the lack of a relationship with group composition, “Where does debate intensity come from?” Several alternative explanations are offered as suggestions for future research.

One alternative source of intensity is the intersubjective responses of the participants (Scheff, 2006). Intersubjectivity suggests that shared cognition and consensus are integral in the creation and evolution of ideas and relations. The participants in any particular DA at any given time react within their own psychological schema while participating in any activity. That is, an individual’s understanding in any given social situation is not the same as anyone else’s in the room. A physician at a DA may be coming from a difficult surgery or an argument with the hospital administrator; a patient at the DA may have been invited because of having written a letter praising the system for its care and concern during a recent visit. When people bring their own psychological development to a situation, it naturally affects the way they view issues, other people, or dialogue among group members. Most of the time, their reactions are unconscious; they are not aware of how much their pasts affect their current behavior and communication with others (Natterson, 1991).

Getting equal representation across a range of stakeholder categories had positive outcomes, but to increase debate, LGI designers may need to also think about the mix of actual personalities from any particular stakeholder group. This suggests that large-group intervention practitioners have an even more complex design choice than simply getting the whole system into the room. Large-group practitioners cannot eliminate the impact of intersubjectivity, but they may be able to set the conditions for intensity, and good facilitators can be aware of group

dynamics affecting the outcome. For example, such instruments as the Myers-Briggs or Big 5 might estimate the coping preferences of potential participants. If the designer is aware of personalities involved in a DA, the facilitators may better understand how intersubjectivity is affecting the group process. As a result, when a particular dialogue becomes increasingly intense or conflictual, each person's internal reaction to information and events happening externally will be different. Hence, because of whom they are, they will respond, behave, and react differently in similar situations that have different participants whether it is due to group dynamics, personality, or anything that can affect intersubjectivity.

Psychological theory regarding intersubjectivity might be helpful for the large-group practitioner. When members come together as a team, "equality of supraordinate goal is a necessary precondition to make their difference a useful complementarity" (Rijsman, 1997). In other words, once participants are defined by personality differences, but also given a common goal (e.g., define an innovative strategy in the DA), "this has the paradoxical character of trying to come closer to each other and then start competing" (Rijsman, 1985). Such a dynamic holds the potential to increase conflict and compliments the intentional designing of the large-group's composition.

A second alternative is that although many contextual and "independent" variables were controlled naturally, the actual content of the discussion was not measured because our focus was on process and composition. Is the development of a cardiology strategy inherently more contentious than a women's and children's strategy? It is possible that the content of the discussion drives disagreement and the question, "What drives a large group to tackle more versus less controversial topics?", is an interesting one. Facilitation is the obvious initial choice. The way the facilitator manages the discussion could generate debate. Similarly, the agenda could be constructed to address more contentious issues and draw out conflict. However, both facilitation and agenda were controlled for here, and so we are left with an intriguing suggestion for future research. (We are grateful to an anonymous reviewer for raising this possibility.)

A third alternative, alluded to in the discussion, is that debate relates to underlying conflict in the interests of the stakeholders and to the expression of power relationships inherent in the system. Despite LGI's normative emphasis on convergence and a positive shared vision of the future, outcomes or processes that challenge traditional power roles may get expressed in overt conflict or in the stifling of emotionally charged discussion. This might explain the

nuanced impact of the overrepresentation or underrepresentation of particular groups as they challenged or succumbed to traditional power roles in surfacing different perspectives and interests.

Habermas (1971) has described how communicative or interpretive knowledge is distinct from technical knowledge. Communicative knowledge facilitates mutual understanding while technical knowledge is based on objective rules and empirical results. He also points out that these two forms of knowledge clash in organizations to create a status hierarchy. In health care, the different disciplinary experts (physicians, technicians, nurses) and administrative staff (managers, housekeeping) are interdependent in the delivery of patient care. Over time, different contributors have achieved high levels of status. Traditional medicine, which seeks clear diagnostic classifications and technologies, has assumed dominance in the institutional order. Even among various physician specialties, there is a distinct hierarchy that relates in part to whether the knowledge base is technical or more diffused and interpretative. For example, surgeons occupy a higher status compared to psychiatrists and specialists see themselves as superior to generalists (Strauss, 1978, Egger and Wagner, 1992).

Thus, a health care system represents a social order that has been constituted through repeated negotiations among professional disciplines with varying degrees of power. Agreements have been reached about what is to be done daily and by whom, what these activities mean, and how they should be justified. This institutional order acts as a cultural constraint on actions, interpretations, and modifications (Strauss et. al., 1964). As people from different professional disciplines work together in an LGI or as part of an innovation network, changes may be suggested that would reconfigure this negotiated order. This may result in conflict and in resistance from the dominant parties that feel their control would be adversely impacted (Manning, 1982). Since formation of a clinical strategy is squarely in the physician's domain, inputs from other stakeholders could be threatening. With a large percentage of physicians in the room, debate could be easily and even unconsciously avoided.

Although LGIs are intended as an empowering, democratic, and high involvement intervention, the presence of high power stakeholders may well require some kind of offset in their design. Those adjustments may include restrictions to powerful stakeholder's involvement (e.g., lowering their representation, confining their contribution to certain times or issues, or

coaching them on the use of their influence). Similarly, it may be necessary to create safe opportunities for less powerful stakeholders to get their perspectives heard.

A second broad suggestion for future research also concerns the debate intensity variable. Given the relatively low debate intensity scores, apparently even small increases were associated with better outcomes. This raises the questions, “if some is good, is more better?” We can imagine that the relationship between debate intensity and outcomes may be curvilinear. That is, increased intensity increases novelty and relevance to a point but too much debate could turn into destructive conflict. LGI’s tendency to focus on common ground may limit debate and constrain the outcomes; but it may also prevent the discussion from getting out of hand. This puts a lot of pressure on the facilitator and supports recent findings by Shmulyian et al. (2010) regarding the importance of that role. Future research should explore the relationships among debate intensity, facilitation style, and LGI outcomes.

Finally, given its central role in the literature, future research should address the influence of the group composition variable. The results of this study suggest that it is balanced representation that generates positive processes and outcomes. The interesting question is how much representation is necessary? In this case, the community category represented a non-homogeneous mix of stakeholders, including customers, patients, regulators, government officials, and so on. But in almost every instance, more of these stakeholders improved processes and results. Future research should explore how variation in the percentage of less central groups has to be present in order to have an impact on the conference’s processes and outcomes.

Conclusion

The six original Alegent Health Right Track DAs represented a rare opportunity to comparatively study large group interventions. A case based literature cannot cumulate knowledge unless one is willing to conduct a meta-analysis of the cases and hope there is enough detail to support the search for patterns in the processes described by others. In the Alegent Health case, we found a naturally occurring field study where six large group interventions occurred within six weeks of each other. It afforded the opportunity to test the assertions of practitioners who have developed these interventions, gave us a chance to validate their claims,

and provided practitioners and organizations with research-based recommendations on how best to leverage these popular change methods.

References

- Axelrod, R. (1992). Getting everyone involved: How one organization involved its employees, supervisors, and managers in redesigning the organization. *Journal of Applied Behavioral Science*, 28, 499-509.
- Baghai, M., Coley, S., & White, D. (2000). *The alchemy of growth*. New York: Basic Books.
- Barrett, F., Thomas, G., & Hocevar, S. (1995). The central role of discourse in large-scale change: A social construction perspective. *Journal of Applied Behavioral Science*, 31, 352-372.
- Beer, M. (2009). *High commitment, high performance*. San Francisco: Jossey-Bass.
- Beer, M., Eisenstat, R. A., & Spector, B. (1990). *The critical path to corporate renewal*. Boston: Harvard Business School Press.
- Bramson, R. A., & Buss, T. (2002). Methods for whole system change in public organizations and communities: An overview of the issues. *Public Organization Review*, 2(3), 211-221.
- Brittain, J., & Freeman, J. (1980). Organizational proliferation and density dependent selection. In J. Kimberly & R. Miles (Eds.) *The Organization Life Cycle*. San Francisco: Jossey-Bass.
- Bryson, J. & Anderson, S. (2000, March/April). Applying large-group interaction methods in the planning and implementation of major change efforts. *Public Administration Review*, 60, 143-162.
- Bunker, B. & Alban, B. (1992a, December). Conclusion: What makes large group interventions effective? *Journal of Applied Behavioral Science*, 28, 579-591.
- Bunker, B. & Alban, B. (1992b, December). The large group intervention—A new social innovation? *Journal of Applied Behavioral Science*, 28, 473-479.

- Bunker, B. B., & Alban, B. T. (1997). *Large group interventions: Engaging the whole system for rapid change*. San Francisco: Jossey-Bass.
- Bunker, B. B., & Alban, B. T. (2005). Introduction to the special issue on large group interventions. *Journal of Applied Behavioral Science*, 41(1), 9-14.
- Cummings, T., & Srivastva, S. (1977). *Management of work*. San Diego: University Associates.
- Cummings, T., & Worley, C. (2009). *Organization development and change* (9th ed.). Mason, OH: Southwestern College Publishing.
- Dannemiller, K., & Jacobs, R. (1992). Changing the way organizations change: A revolution of common sense. *The Journal of Applied Behavioral Science*, 28, 480-498.
- Egger, E. and I. Wagner. (1992). Time management: A case for CSCW. Proceedings of the 1992 ACM conference on computer-supported cooperative work, pp. 249-256. Toronto, Canada
- Fagerberg, J. (2004). Innovation: A guide to the literature. In J. Fagerberg, D. Mowery, & R. Nelson (Eds.), *The Oxford handbook of innovations* (pp. 1-26). New York: Oxford University Press.
- Fredrickson, J., & Mitchell, T. (1984). Strategic decision processes: Comprehensiveness and performance in an industry with an unstable environment. *Academy of Management Journal*. 27: 399-424.
- Garcia, S. (2007). Developing social network propositions to explain large-group intervention theory and practice. *Advances in Developing Human Resources*, 9, 341-358.
- Habermas, J. (1971). *Knowledge and human interests*. Boston: Beacon Press.
- Hamel, G. (2007). *The future of management*. Boston: Harvard Business School Press.

- Holman, P., T. Devane, and S. Cady. (2007). *The Change Handbook*. San Francisco: Berrett-Koehler.
- Kotter, J. (1996). *Leading change*. Boston: Harvard Business School Press.
- Lawler, E. (1986). *High-involvement management*. San Francisco: Jossey-Bass.
- Lawler, E., & Worley, C. (2006). *Built to change*. San Francisco: Jossey-Bass.
- Ledford, G. (1984). *The persistence of planned organizational change: A process theory perspective*. Doctoral dissertation, University of Michigan.
- Manning, P. (1982). Organizational work: Structuration of environments, *British Journal of Sociology*, 33, 118-134.
- Manning, M. R., & Binzagr, G. F. (1996, July). Methods, values, and assumptions underlying large group interventions intended to change whole systems. *Journal of Organizational Analysis*, 4, 268-284.
- Miller, M., Fitzgerald, S., Murrell, K., Preston, J., & Ambekar, R. (2005). Appreciative inquiry in building a transcultural strategic alliance. *Journal of Applied Behavioral Science*, 41, 91-110.
- Natterson, J. (1991). *Beyond countertransference: The therapist's subjectivity in the therapeutic process*. New Jersey: Joseph Aronson, Inc.
- Owen, H. (1992). *Open space technology*. Potomac, MD: Abbott.
- Pettigrew, A., & Fenton, E. (2000). *The innovating organization*. Thousand Oaks, CA: Sage Publications.
- Purser, R., & Griffin, T. (2008). Large group interventions: Whole system approaches to organizational change. In T. Cummings (Ed.), *Handbook of Organization Development and Change* (pp. 261-276). Thousand Oaks, CA: Sage Publications.

- Rijsman, J. (1985). Some equalitarian aspects of social competition. In F. Denmark (Ed.), *Socioecological psychology and the psychology of women* (pp. 167-176). Amsterdam: North Holland.
- Rijsman, J. (1997). Social diversity: A social psychological analysis and some implications for groups and organizations. *European Journal of Work and Organizational Psychology*, 6 (2): 139-152.
- Scheff, T. (2006). *Goffman unbound: A new paradigm for social science*. New York: Paradigm Publishers.
- Shmulyian, S., B. Bateman, R. Philpott, and N. Gulri. (2010). Art or artist? An analysis of eight large-group methods for driving large-scale change. In W. Pasmore, A. Shani, and R. Woodman (Eds.), *Research in Organizational Change and Development* vol. 18 (pp. 183-232).
- Strauss, A. (1978). *Negotiations: Varieties, context, processes, and social order*. San Francisco: Jossey-Bass.
- Strauss, A., L. Schatzman, R. Bucher, D. Erlich, and M. Sabshin. (1964). *Psychiatric ideologies and institutions*. New York: Free Press.
- Tenkasi, R., Mohrman, S., & Mohrman, A. (1998). Accelerated learning during organizational transition. In S. Mohrman, J. Galbraith, E. Lawler & Associates (Eds.). *Tomorrow's organization*. San Francisco: Jossey-Bass.
- Weisbord, M. (1987). *Productive workplaces*. San Francisco: Jossey-Bass.
- Weisbord, M. R. (1992). *Discovering common ground*. San Francisco: Berrett-Koehler.
- Weisbord, M., & Janoff, S. (2005). Faster, shorter, cheaper may be simple; It's never easy. *Journal of Applied Behavioral Science*, 41, 70-82.

Wheatley, M. (1992). *Leadership and the new science*. San Francisco: Berrett-Koehler.

Worley, C., & Feyerherm, A. (2003). Reflections on the future of OD. *Journal of Applied Behavioral Science*, 39 (1): 97-115.

Appendix A – Interview Questions for Executives

- Tell me a little about your background and history with the health care system:
- What do you see as the strengths of this organization? What weaknesses do you worry about?
- We'd like to hear your observations and opinions about the health care system's strategy and organization.
 - How would you characterize the organization's business and clinical strategies two years ago?
 - How would you characterize them today?
 - How have they changed? Why?
 - What, for you, have been the big organizational changes in the system?
 - Probe for structural, systems, and process changes at each level
- To what extent do you believe these changes in strategy and structure are attributable to the [large group intervention] process? What other forces or decisions are responsible for these changes?
- Thinking about the original six "clinical strategy" DAs, how many of them did you attend or participate it?
 - Please take a few minutes to complete this survey – It asks for your opinions about the service line strategies you are most familiar with.
 - For any DA event that you are familiar with, what were the biggest tensions during the process?
 - What have been the biggest tensions in the process overall?
- To what extent are people in the organization aware of the [large group intervention] program and process? How do they hear about it? Which method has proved most effective?
- From your point of view, how clear were the action plans and other implementation activities coming out of the initial Right Track service line workshops?
- What is going well/not so well in implementation of the different service line strategies?
- What would you list as the "high points" or "best success stories" that have resulted because of the RT process (not the RT process itself)?
- What have been some of the limitations or problems caused by the RT process? Why?
- How well is the facilities/capital improvement process supporting the service line strategies?
- How well are the other enabling strategies—e.g., IT, Intensive Care, etc.—supporting the service line strategies?
- Overall, how do you feel about the contribution of the DA process to date, and the manner in which Alegend is using it?
- What changes do you think should be made in order to improve the value that Alegend is getting from the RT process?
- For us to more fully understand the strategy process and the role that RT has played—and how the organization is translating that work into actual change in the way Alegend operates, are there key people we should talk to?
 - Leaders of the strategy and implementation activities?
 - Leaders of the facilities that are impacted?
 - Key opinion leaders?