Deliberate Learning to Improve Performance in Dynamic Service Settings: Evidence from Hospital Intensive Care Units

Ingrid M. Nemhhard
Yale University, School of Medicine and School of Management, New Haven, Connecticut 06520, ingrid.nembhard@yale.edu

Anita L. Tucker
Harvard Business School, Boston, Massachusetts 02163, atucker@hbs.edu

Dynamic service settings—characterized by workers who interact with customers to deliver services in a rapidly changing, uncertain, and complex environment (e.g., hospitals)—play an important role in the economy. Organizational learning studies in these settings have largely investigated autonomous learning via cumulative experience as a strategy for performance improvement. Whether induced learning through the use of deliberate learning activities provides additional performance benefits has been neglected. We argue that the use of deliberate learning activities offers performance benefits beyond those of cumulative experience because these activities counter the learning challenges presented by rapid knowledge growth, uncertainty, and complexity in dynamic settings. We test whether there are additional performance benefits to using deliberate learning activities and whether the effectiveness of these activities depends on interdisciplinary collaboration in the workgroup. We test our hypotheses in a study of 23 hospital neonatal intensive care units (NICUs) involved in a quality improvement collaborative. We find that using deliberate learning activities is associated with better workgroup performance, as measured by NICUs’ risk-adjusted mortality rates for 2159 infant patients, but only after two years. In the shorter term, using these activities is associated with worse performance. By the third year, the positive impact of using deliberate learning activities is similar to the benefit of cumulative experience (18% and 20% reduction in odds of mortality, respectively). Contrary to prediction, interdisciplinary collaboration mediates, rather than moderates, the relationship between using deliberate learning activities and workgroup performance. Thus, our data suggest that using deliberate learning activities fosters interdisciplinary collaboration.

Key words: organizational learning; deliberate learning; cumulative experience; induced learning; autonomous learning; interdisciplinary collaboration; health care; hospitals

History: Published online in Articles in Advance September 15, 2010.

1. Introduction
Organizational learning—the process of improving organizational actions by integrating new knowledge and insights—has been extensively discussed in the management literature (Argote 1999, Fiol and Lyles 1985, Levitt and March 1988). The interest stems from the belief that organizational learning is the key to success in dynamic environments (Edmondson 2008, Senge 1990), a characterization that has come to describe many industries (Cascio 2003).

Scholars assert that organizations can learn either through autonomous learning or induced learning (Dutton and Thomas 1984). Autonomous learning refers to incidental learning that occurs as employees accumulate experience through task repetition. This type of learning is tacit and captured in the notions of “learning-by-doing” and “first-order learning” (Adler and Clark 1991). In contrast to autonomous learning, induced learning refers to learning that stems from deliberate activities to create, acquire, or transfer knowledge. Examples include training, experiments, dry runs of practices, quality management programs, and suggestion programs (Adler and Clark 1991, Arthur and Huntley 2005, Laprê et al. 2000, Tucker et al. 2007). This more cognitive approach has been labeled “planned learning” (Levy 1965), “second-order learning” (Adler and Clark 1991), and “deliberate learning” (Zollo and Winter 2002). This approach assumes that choices made by managers and staff change the rate at which learning occurs.

An extensive body of empirical research has documented a positive relationship between autonomous learning as measured by cumulative experience (e.g., cumulative volume and calendar time) and various measures of performance improvement (e.g., productivity increases, cost reductions, higher yield rates, etc.). (See Argote 1999 for a review of the research.) This body of research also has documented considerable variation in the rate at which performance improves with experience. Dutton and Thomas (1984) reviewed more than 100 studies and found that learning rates in manufacturing
settings varied from 55% to 108%. Studies in dynamic service settings show similar variation in organizational learning from cumulative experience (Darr et al. 1995, Lapré and Tsikriktsis 2006, Reagans et al. 2005).

Variation in learning rates has sparked interest in induced learning (Arthur and Huntley 2005, Lapré and Van Wassenhove 2001, Zollo and Winter 2002) and has raised an important theoretical and empirical question: To what extent is it possible for organizations to use induced learning to enhance their performance beyond the benefits derived from autonomous learning? Or, as we ultimately consider, does induced learning as measured by the use of deliberate learning activities (DLAs) offer performance benefits beyond those gained from autonomous learning as measured by cumulative experience? This question is relevant across work settings, but there is an imperative to answer it for dynamic service settings for two reasons. First, dynamic service settings, which are characterized by workers who interact with clients to deliver services in a rapidly changing, uncertain, and complex environment, play an important role in our economy. The service sector accounts for 70% of the U.S. economy and provides more than three out of four jobs (U.S. Bureau of Economic Analysis 2009). Second, evidence of poor performance in dynamic service settings is abundant. According to the National Quality Research Center (2009), the American Customer Satisfaction Index in service settings ranges from 64% to 79%.

To our knowledge, with the exception of research by Pisano et al. (2001), there is little empirical research that addresses whether induced learning (as measured by DLA) offers performance benefits beyond those gained from autonomous learning (as measured by cumulative experience) in dynamic service settings. Pisano et al. (2001) studied organizational differences in rates of learning how to perform minimally invasive cardiac surgery in 16 hospitals. They found considerable variation in learning rates, with the best performer able to complete the surgery in 143 minutes after 40 cases, whereas the worst performer required 305 minutes. The researchers proposed that the difference in learning rates reflected differential use of DLA (i.e., practice session and early trials), a proposition they derived from qualitative case study analyses but were unable to test formally because they did not measure the use of DLA.

This paper aims to further the understanding of organizational learning in dynamic service settings by assessing the effectiveness of induced learning on workgroup performance, after accounting for the impact of autonomous learning in such settings. In our discussion, we focus on measurable forms of induced and autonomous learning: “deliberate learning activities” and “cumulative experience,” respectively. We focus on workgroup performance because an emerging consensus holds that organizational learning occurs through the actions and interactions of individuals working in small workgroups or teams (Edmondson 2002, Senge 1990). A workgroup is a group of individuals that exists within a larger organization, has a clearly defined membership, and is responsible for a shared product or service (Hackman 1987). In the context of a hospital, for example, workgroups are the multidisciplinary groups of professionals that deliver a specific domain of clinical care.

The structure of our paper is as follows. We begin with a description of our research setting, health-care delivery, to provide an example of a dynamic work setting and explain why organizational learning is important in this setting. We then review the organizational learning literature to develop hypotheses about how using DLA affects workgroup performance given the workgroup’s cumulative experience and how the effectiveness of DLA is altered by a critical interaction in workgroups, interdisciplinary collaboration. Our hypotheses draw on the organizational learning literature from two historically disparate disciplines: organizational behavior and operations management. We test our hypotheses in a three-year study of 23 hospital neonatal intensive care units (NICUs). We then report results related to our measure of workgroup performance, NICUs’ risk-adjusted mortality rates, for 2,159 infant patients. The ability to assess the relationship between using DLA, a managerial intervention, and patient mortality is a unique advantage of this study. We conclude with a discussion of the implications of our findings for future research and managerial practice in health care and other dynamic service settings.

2. Research Setting: The Dynamic Health-Care Environment

Organizational learning in dynamic service settings poses a great challenge because of three characteristics of these settings: rapid change, complexity, and uncertainty. Although many settings (e.g., emergency response and professional services such as investment services) share these characteristics, health-care delivery—the setting for our study—is a classic example of a dynamic service setting and one where organizations vary in their use of autonomous and induced learning.

Health-Care Delivery as an Example. Health-care professionals (physicians, nurses, therapists, etc.) face an astonishing rate of change in medical knowledge as a result of rapid advances in science and technology. By some estimates, the number of health-care–related randomized control trials—the gold standard for evidence of new knowledge in medicine—increases at a rate of 10,000 annually (Institute of Medicine 1999). Knowledge derived from other forms of research is growing as well, as evidenced by the addition of 30,000 new references each month (on average) to the MEDLINE bibliographic database. The rapid increase in knowledge about more effective and efficient ways to provide care means
health-care organizations and professionals find themselves continuously challenged to learn new practices. Even though the medical knowledge base is large and growing, health-care professionals still face much uncertainty in their work (Bohmer 2009). Health-care delivery is not an exact science because disease manifests itself differently across patients. This creates variability in “customers”—a key characteristic of service settings. Customer variability leads to uncertainty. For example, there is often uncertainty about patient diagnosis and the best treatment. Multiple treatment options often exist, and the risks and benefits of each treatment for an individual patient are unknown (Nembhard et al. 2009). Finally, patient conditions contribute to uncertainty because they can change unpredictably during the course of treatment. Thus health-care professionals often have to adjust in real time, a function of the simultaneity of production and consumption in service settings. Simultaneity, the need to work with the customer (e.g., patient) to deliver the service, and customer variability are hallmarks of service settings that distinguish them from manufacturing settings (Bitran and Lojo 1993). The fact that different types of patients arrive emergently adds another layer of uncertainty and dynamism to this setting (Argote 1982).

Adding to the dynamism, health-care professionals encounter a high degree of complexity at work. Driven largely by the enormous knowledge base, the workforce has become highly specialized, and the number of specialties continues to increase at a rapid rate (Nembhard et al. 2009). The high degree of specialization means that each professional brings only a fraction of the knowledge needed to care for patients. By some estimates, more than 20 health-care professionals (multiple types of physicians, nurses, therapists, etc.) must integrate their expertise to provide care for a single patient in a hospital (Bohmer and Knoop 2007). Thus there is a high level of interdependence in this setting, and the interdependence continues to grow because new technologies and care practices increasingly involve reciprocal (as opposed to sequential) processes. The huck-and-forth nature of the work means that health-care professionals cannot implement new knowledge without collaborating across disciplines. Nevertheless, interdisciplinary collaboration is often missing in workgroup interactions (Institute of Medicine 2004).

A series of seminal reports have shown that health-care organizations have struggled to perform well under these dynamic conditions with respect to a key performance metric: quality of care. For example, a RAND study showed that only 55% of patients receive the recommended care for their condition from their care providers (McGlynn et al. 2003). Additionally, nearly 2.5 million patients unnecessarily suffer serious, adverse events such as hospital-acquired infection or wrong-limb amputation each year (Leape 1997), and nearly 100,000 preventable deaths caused by medical errors occur each year, the majority stemming from collaboration failures (Institute of Medicine 1999).

Induced vs. Autonomous Learning as a Solution. Health-care organizations have differed in their attempts to address quality problems, with some emphasizing induced learning and others emphasizing autonomous learning. Virginia Mason Medical Center (VMMC) in Seattle, Washington, and Shouldice Hospital in Ontario, Canada, for example, have taken opposite approaches. VMMC has adapted the Toyota Production System and its principles of lean manufacturing as its strategy (Bohmer 2009). At VMMC, workgroups are charged with proactively addressing problems. In addition to having the authority to “stop the line” when problems occur to conduct root cause analysis with managers and quickly implement solutions, individuals and workgroups at VMMC continually experiment with ways to improve their practices. In other words, VMMC focuses on induced learning through DLA to improve quality of care. In contrast, Shouldice Hospital relies on autonomous learning through cumulative experience (Heskett 2003). It operates as a high-volume, specialty hospital that only treats patients that require hernia operations, and it aims to improve care through focused operations and task repetition.

Shouldice’s approach reflects a prominent view in health care that organizations can achieve high-quality care if they accumulate sufficient experience. This belief has led some to recommend mandating minimum levels of patient volume. For example, guidelines from The Leapfrog Group, an employer-action group formed to improve health-care quality, stipulate that a hospital must treat at least 450 patients annually for coronary bypass graft surgery to receive reward and recognition (Lwin and Shepard 2008). These recommendations are rooted in research showing a positive association between patient volume and clinical performance (e.g., lower mortality rates). (See Halm et al. 2002 for a comprehensive review of these “volume-outcome studies.”) Notably, these studies largely examined performance for procedures that had been in place for several years. They did not examine performance under the dynamic conditions more characteristic of health-care delivery today. Moreover, research now has shown variability in learning rates based on volume in health care (Pisano et al. 2001, Reagans et al. 2005), suggesting increased volume is not a sufficient strategy for quality improvement in health care. In this paper, we examine the potential added contribution of DLA to performance in health care, an important enterprise given the need for quality improvement and the slow pace of progress in this setting (Kuehn 2009).
3. Theory and Hypotheses

3.1. The Role of Induced Learning in Dynamic Service Settings

We argue that induced learning through the use of DLA is an effective strategy for countering the challenges of rapid knowledge growth, uncertainty, and complexity present in dynamic settings. Therefore, these activities should offer performance benefits beyond those provided by cumulative experience. Rapid knowledge growth, for example, means workgroups constantly have new practices to learn. Many of these new practices dramatically alter existing workgroup routines or make existing routines obsolete (e.g., the shift from traditional, open-heart cardiac surgery to minimally invasive cardiac surgery; Pisano et al. 2001). Moreover, newer practices in dynamic service settings like health care tend to be more complex and contain a large tacit component (Berta and Baker 2004). These characteristics limit workgroups’ ability to draw upon their past experiences to learn new practices, thus limiting workgroups’ absorptive capacity for learning (Cohen and Levinthal 1990).

Using DLA may make it easier for workgroups to absorb new practices because these activities foster a deep understanding of new practices. In a study of 62 quality improvement projects, Mukherjee et al. (1998) found that using DLA can generate both operational knowledge about how to perform new practices effectively and conceptual knowledge about cause-and-effect relationships that make practices effective. Together, these two types of knowledge increased project teams’ ability to alter routines and achieve their goals. Thus we expect the use of DLA to enhance performance in dynamic service settings by increasing workgroups’ absorptive capacity for learning new practices that are a significant departure from existing routines.

The use of DLA should also help counter the challenges to learning posed by uncertainty in dynamic service settings, particularly those derived from customer variability and the simultaneity of production and consumption. Customer variability can limit opportunities to develop skill with new practices through experience. This is because practices may only be relevant to a specific customer segment that may not be increased easily to gain skill-building experience quickly. In many hospitals, for example, gaining expertise in the treatment of patients with a type of heart attack for which mortality increases significantly if not treated within 90 minutes (ST-segment elevation myocardial infarction) is challenged by having only two cases on average each month (McNamara et al. 2006). Levitt and March (1988, p. 333) refer to this challenge as the “the paucity of experience problem.” They argue that “learning from experience in organizations is compromised by the fact that nature provides inadequate experience relative to the complexities and instabilities particularly when the environment is changing rapidly.” DLA provide alternative opportunities for workgroups to gain expertise by “learning-before-doing” (Upton and Kim 1998) so that workgroups may be prepared even for their first service encounter. Thus we expect the use of these activities to provide benefits by virtue of the opportunity they provide for added learning.

Opportunity is important, but insufficient, if workgroups are not motivated to learn (Argote et al. 2003). Motivation to learn is often limited in dynamic service settings because the simultaneity of production and consumption in these settings means that mishaps in the course of learning may be felt by customers. Research suggests that many health-care professionals avoid implementing new practices because they fear causing harm to patients (Nembhard et al. 2009). The use of DLA creates a relatively low-risk forum in which workgroups have protected time and opportunity to learn new practices and how to manage the interdependencies brought about by new practices. As workgroups become savvy in managing the interdependencies, uncertainty and felt complexity should be reduced, enabling workgroups to perform at a higher level. Thus we propose that DLA enhance workgroup performance because they help workgroups overcome challenges presented by knowledge growth, uncertainty, and complexity.

Empirical evidence from learning studies of manufacturing firms developing new production processes seem to support our supposition that DLA (measured by investment in engineering activities in these studies) provide additional benefits. For example, Hatch and Mowery (1998) found that improvement in yield rates in the semiconductor industry were a function of investment in engineering analysis but not of cumulative volume in the earliest stages of developing new manufacturing processes. The influence of both factors equalized as processes matured. The study of electronic component manufacturers by Ittner et al. (2001) also showed that past cumulative volume and past cumulative expenditure on conformance activities contributed equally to firms’ current quality levels. These results suggest that DLA can offer performance benefits beyond cumulative experience in dynamic settings. Thus, we hypothesize the following.

Hypothesis 1. The use of deliberate learning activities is positively associated with workgroup performance in dynamic service settings, beyond the positive contribution of cumulative experience.

3.2. The Moderating Role of Interdisciplinary Collaboration

In dynamic work settings, performance benefits derived from the use of DLA likely depend upon interactions within the workgroup, given the interdependence among professionals in these settings. Specifically, the benefits
likely depend on the level of interdisciplinary collaboration in the workgroup. Interdisciplinary collaboration refers to the degree of cooperation among individuals with different disciplinary backgrounds (Baggs et al. 1999, Jassawalla and Sashittal 1998, Pinto et al. 1993). The literature on interdisciplinary groups and teams suggests three reasons why greater interdisciplinary collaboration may improve the effectiveness of DLA in dynamic work settings.

The first relates to the quality of decisions made about when, how, and where to use DLA. Better decisions about routine activities are made when professionals from different disciplines collaborate. Collaborators elevate the quality of the decision-making process by openly sharing their expertise, raising questions, fully considering alternatives, and integrating ideas across disciplines (Simons et al. 1999). We expect that workgroups that have a norm of interdisciplinary collaboration for decision making about routine activities will also collaborate about DLA, leading to higher-quality decisions about their use and better integration of knowledge gained from their use into other decisions that impact performance.

The second reason for why it is likely that DLA will be more effective in the context of interdisciplinary collaboration is that proper execution of DLA in dynamic work settings typically requires coordination, i.e., the management and integration of different pieces of the task to accomplish the collective goal (Van de Ven et al. 1976). The need for coordination largely reflects the reciprocal interdependence resulting from task complexity in these settings. Research has found that collaborators develop transactive memory about “who knows what” (Liang et al. 1995, Reagans et al. 2005). That knowledge facilitates the development of a shared understanding that enables collaborators to coordinate their work effectively (Gittell 2002). Liang et al. (1995) showed that collaborators experienced less need for planning, less confusion, and fewer misunderstandings. Gittell (2002) further showed that interdisciplinary collaborators accomplished their work more efficiently and at higher quality. Building on these findings, we expect that using DLA in workgroups with interdisciplinary collaboration will yield better results because these workgroups will effectively coordinate their efforts during these activities and therefore maximize the benefits. In contrast, we expect workgroups with low levels of interdisciplinary collaboration to struggle with coordination, reducing the effectiveness of their DLA.

The third reason we expect higher levels of interdisciplinary collaboration to improve the effectiveness of DLA is because collaborators are skilled at error detection and recovery. Jassawalla and Sashittal (1998) observed that collaborators in the new product development process displayed collective mindfulness, i.e., a shared vigilance to the task and each other. Such mindfulness increases the detection of changing conditions and errors, enabling collaborators to respond sooner to minimize adverse effects on performance (Weick et al. 1999). We posit that collaborators’ superior skill for error detection and recovery is likely to be beneficial when they are engaged in DLA because errors are natural occurrences during these activities. To the extent collaborators detect and learn from errors more effectively, the workgroup’s ability to benefit from DLA should be enhanced. Given the multiple ways in which norms of interdisciplinary collaboration within the workgroup complement and facilitate the use of DLA, we hypothesize a moderating effect.

Hypothesis 2. The positive relationship between deliberate learning activities and workgroup performance is stronger in workgroups with greater interdisciplinary collaboration.

4. Methods

4.1. Research Setting

We conducted this research in collaboration with the Vermont Oxford Network (VON), a professional association for NICUs. NICUs are hospital-based units that provide care for premature and critically ill infants. Each year, approximately 8% of the 4.1 million infants born in the United States are born prematurely or with low birth weight and therefore are admitted to an NICU (Martin et al. 2007). In the NICU, infants receive care from a multidisciplinary group of professionals, including neonatologists (physicians that provide care to neonates), nurses, respiratory therapists, pharmacists, and nutritionists. As a group, these professionals face multiple challenges. Every day, they must deliver care that can be very complex, particularly for the sickest infants. Additionally, because the neonatal knowledge base is rapidly evolving to care for younger and sicker infants, neonatal health professionals constantly face the challenge of implementing new, and often more complex, work practices.

In 2002, 44 VON-member NICUs from the United States and Canada formed a “quality improvement collaborative” to work together to implement new work practices, making this collaborative an ideal setting for an empirical test of our hypotheses about DLA. In the collaborative, each NICU was represented by an improvement project team. The teams—typically consisting of at least one physician, a nurse, a respiratory therapist, and the unit manager—attended collaborative meetings every six months for two years. At the meetings, experts taught them strategies for effectively changing work processes, such as rapid cycle improvement, root cause analysis, and systems thinking. The teams also used these meetings to identify potentially better
practices (PBPs); i.e., practices that the medical literature suggested could, but were not guaranteed to, lead to better outcomes (Horbar et al. 2001). Working together, the teams identified a total of 93 PBPs across seven target areas, including for example, infection control, respiratory care, and discharge planning. Each NICU decided which subset of the 93 PBPs it wanted to implement. On average, each selected 36.2 PBPs. Thus, there was ample opportunity to assess the use of DLA and gather data on its relationships to interdisciplinary collaboration and workgroup performance.

4.2. Data Collection
Our data came from two sources. Our first source was a survey we administered to staff in the NICUs in 2003. The survey contained questions that had been asked of staff in previous collaboratives (Baker et al. 2003). It also included questions we developed to assess additional concepts present in the organizational learning literature and discussed with us during preliminary interviews at four of the NICUs. To ensure our questions were conceptually and psychometrically valid, we piloted the survey in four sites before inviting the remaining 40 NICUs in the collaborative to participate in our survey. The survey had two parts. Part I, which we asked all staff to complete, included questions about interdisciplinary collaboration. Part II, which only improvement project team members completed, included questions about DLA.

Twenty-three of the 40 NICUs agreed to participate in our survey for an NICU response rate of 58%. These NICUs did not differ significantly from nonparticipants on a variety of structural and clinical measures (e.g., hospital ownership, teaching status, patient acuity, and the number of past collaboratives in which the NICU had participated), which minimized our concern about response bias. Across the participating NICUs, we administered the survey to 3,130 people and received 1,440 responses, for an individual response rate of 46%. Our respondents, a mean of 63 per NICU, represented a variety of disciplines. On average, 49 respondents not involved with the improvement projects (referred to as “nonimprovement project staff”) completed Part I of the survey. Part II was completed by an average of 11 people per NICU (minimum = 3, maximum = 35). The total sample size for Part II was 265 respondents.

Our second data source was an NICU patient database maintained by VON. This database contained demographic information on each NICU (e.g., teaching status) that helped us characterize our sample. It also contained demographic and outcome data on every infant who was a patient at a VON member NICU in 2001, 2003, and 2004. We tested our hypotheses with data on infants born during 2003 and 2004, the years after the NICUs began using DLA to implement the PBPs. We used the 2001 data to adjust for performance prior to the start of the collaborative. We obtained data on a total of 1,040 infant patients for 2001 (mean = 45.2, S.D. = 22.4), 1,102 infant patients in 2003 (mean = 47.9, S.D. = 26), and 1,072 infant patients for 2004 (mean = 46.8, S.D. = 27.8). All of these infants had an extremely low birth weight (ELBW), weighing less than 1,000 grams, or 2.2 pounds, at birth. Infants with higher birth weight, severe birth defects, or who did not live at least three days were excluded from our sample. Additionally, we excluded infants for whom data were missing, leading to a slightly smaller sample size of 1,098 infants for 2003 and 1,061 for 2004, a total 2,159 infants in our data set for analysis.

4.3. Measures
Use of deliberate learning activities (DLA) was assessed using improvement project teams’ survey ratings of how frequently their team used each of seven learning activities during the course of their improvement project (1 = not at all, 5 = to a very great extent). The activities included solicitation of staff ideas, opportunities for staff to provide feedback before full implementation, education sessions with staff, pilot runs, dry runs, project team meetings, and problem-solving cycles (Plan-Do-Study-Act). This bundle of activities was called “learn-how” in a previous paper using this data set because the study teams were learning how to implement new practices through these activities (Tucker et al. 2007). In this paper we call it “deliberate learning activities” to better describe the purpose of the activities. Cronbach’s alpha for this scale was 0.88. Because there was a high level of agreement among the ratings of project team members from the same NICU (mean $r_{WG} = 0.61$), we aggregated their responses to the NICU level by calculating the arithmetic mean. Differences in ratings between NICUs also supported aggregation: intraclass correlation coefficient ICC(1), indicating the ratio of NICU-level variance to total variance, was 0.11 (analysis of variance (ANOVA) $F = 1.83$, $p = 0.01$); and ICC(2), a measure of the reliability of average NICU perceptions, was 0.45.

Cumulative experience was measured by cumulative patient volume during the three-year collaborative period. Specifically, for the 2003 ELBW infants, we summed the number of ELBW infants treated in the NICU during 2002 and 2003; for the 2004 infants, we summed the number of babies in 2002, 2003, and 2004. Because we did not have data from 2002, we estimated the infants treated that year using the mean number of babies treated in 2001, 2003, and 2004. We only included experience acquired during the collaborative because membership in the collaborative arguably represented a new starting point in the NICUs’ experience, and research also shows that older experience has less value (Argote 1999).

Interdisciplinary collaboration was assessed using the reported level of agreement (1 = strongly disagree, 7 =
strongly agree) with six survey items completed by non-improvement project staff who were not physicians (e.g., nurses and respiratory therapists). 1 Four of these items (e.g., “Communication between nurses and physicians is open and positive”) were adapted from a scale used by Baker et al. (2003). The two remaining items (e.g., “The input of respiratory therapists and/or other ancillary staff is regularly sought when developing treatment plans”) were developed by us to broaden the range of disciplines considered. Cronbach’s alpha for the six-item scale was 0.88. As with project team members’ responses to items about DLA, we averaged responses within each NICU to arrive at NICU-level measures of interdisciplinary collaboration. Both the consistency in responses among staff within NICUs (mean $r_{wg} = 0.75$) and the differences in assessment between NICUs (ICC(1) = 0.33, ANOVA $F = 15.12, p < 0.001$; ICC(2) = 0.93) supported aggregation. All survey items are shown in the appendix.

Workgroup performance was assessed using an established measure of success for hospital units: patient mortality (Mitchell and Shortell 1997, West et al. 2006). We used an indicator variable to denote whether each infant treated in our sample died while in the NICU. We focused on the mortality of infants born in 2003 and 2004 because the collaborative effort was well underway during those years, and it was expected that the performance effects of using DLA would be observable. Prior research on collaboratives has shown significant performance effects after two years (e.g., Horbar et al. 2001).

Control Variables. We included two sets of control variables in our analysis. The first set controlled for infant characteristics known to influence mortality, namely, those in the risk-adjustment model developed by the Vermont Oxford Network (Rogowski et al. 2004): gestational age, gestational age squared, gender (1 = male, 0 = female), small for gestational age (1 = birth weight less than the 10th percentile for gestational age based on race and sex, 0 = otherwise), member of multiple birth (e.g., twin; 1 = yes, 0 = no), vaginal delivery (1 = yes, 0 = no), outborn status (1 = infant born at another hospital, 0 = infant born at current hospital), mother’s race (set of dummy variables with black as the reference), and one-minute Apgar score (ranges from 0 to 10, with higher values indicating better health based on five indicators: heart rate, color, reflexive grimacing, breathing, and activity). Research shows that the VON risk-adjustment model works well in predicting infant mortality, with an area under the operating curve of 0.88 (Rogowski et al. 2004).

The second set of variables controlled for NICU characteristics outside of our model that could have influenced patient outcomes. With the small number of NICUs in this study, we could not include all potential NICU-level control variables. Therefore, we added control variables one at a time to our model and retained those that were statistically significant ($p < 0.10$) (Becker 2005): urban location (1 = urban, 0 = otherwise), teaching status (1 = teaching occurs in NICU, 0 = otherwise), improvement team size, and average staff tenure, which was our proxy for staff turnover. Staff tenure in the unit was measured as a categorical variable with 1 = less than one year working in this NICU, 2 = more than one year but less than two years, 3 = between two and five years, 4 = between five and ten years, and 5 = ten years or more. Although this is an imperfect measure of staff turnover, it was the best available because we were unable to obtain measures of turnover from each NICU. The following potential control variables were tried but found to be insignificant: hospital ownership status (e.g., for-profit hospital), number of NICU beds, staff-to-beds ratio, average staff hours worked weekly, percentage of physicians on the improvement team, number of times the NICU participated in a VON collaborative, number of improvement projects during the collaborative, and level of evidence supporting the NICU’s improvement project portfolio. Psychological safety—a measure of culture that assesses the extent to which staff feel it is safe to speak up with questions, concerns, and suggestions—was also found to be an insignificant predictor of patient mortality; therefore, we excluded it from tests of our hypotheses. However, we included this measure as a control variable in supplemental analyses examining alternative models of the relationships between DLA, collaboration, and patient mortality. We included it because prior research found a link between psychological safety and DLA (Tucker et al. 2007). We used the Tucker et al. (2007) three-item scale of psychological safety (shown in the appendix; Cronbach’s alpha $= 0.74$) in our supplemental analyses.

To facilitate interpretation of directional effects, we controlled for prior workgroup performance. Following West et al. (2006), we included each NICU’s 2001 risk-adjusted standardized mortality ratio (SMR) as a control variable. We used the SMR for 2001 because 2001 is the year before the NICUs began their DLA as members of the collaborative. To calculate the SMR for each NICU, we first computed the predicted risk-adjusted mortality for each ELBW infant born in that NICU during 2001 using the VON risk adjustment model (Rogowski et al. 2004). We used logistic regression with clustering of infants by NICU and with the between-NICU variation modeled as random effects to predict the probability of death for each infant, given the infant risk characteristics mentioned above. For each NICU’s infants born in 2001, we summed these predictions and created a ratio of predicted death over the sum of actual deaths. This ratio is the SMR. SMRs less than one indicate better workgroup performance (less mortality) than expected, whereas SMRs greater than one indicate worse workgroup performance.

4.4. Data Analysis
We calculated descriptive statistics for the 2,159 infants and 23 NICUs in our sample. We then conducted
correlation analysis to examine the bivariate relationships between DLA, cumulative experience, interdisciplinary collaboration, and mortality. However, this analysis was an insufficient test of our hypotheses because patient mortality, our measure of workgroup performance, required risk adjustment for patient characteristics before its relationship to other variables could be assessed meaningfully (Kahn et al. 2006). Thus we formally tested our hypotheses with multivariate regression analyses that allowed us to adjust for the characteristics of infants in our sample.

We conducted our analyses with workgroup performance (patient mortality) measured at the patient level \((N = 2,159)\), a common approach in health services research because patient-level data allow the researcher to leverage the larger sample of patients to identify significant unit- and hospital-level effects on patient outcomes (Kahn et al. 2006). We used clustered variance estimators with our regression models to account for correlation among infants treated in the same NICU. We used logistic regression models because our dependent variable, patient mortality, was binary. In logistic regression models, odds ratios (ORs) indicate the relationship between variables. In this study, an OR of less than one indicates that an increase in the level of a predictor (e.g., use of DLA) is associated with better workgroup performance because it indicates that an infant treated in the NICU is less likely to die when the level of the predictor increases. Conversely, an OR greater than one indicates that an increase in the level of a predictor is associated with worse workgroup performance (i.e., higher mortality). We standardized our independent variables to facilitate comparison of effect sizes.

We assessed support for Hypothesis 1 in two ways. First, we examined the ORs for our variables of interest—the use of DLA and cumulative experience—in a regression across all infants in our sample. We did not differentiate between infants born in 2003 and 2004. In our second analysis, we examined the effect of our variables on infants born in 2003 versus those born in 2004 to assess whether the impact of the variables, and thus support for our hypothesis, changed over time. To do this, we created two dummy variables for birth year—one indicated whether the infant was born in 2003, and the second whether the infant was born in 2004. We multiplied our two variables of interest by the dummy variables to create new variables representing the impact of the variables in each year. We included the new variables and the dummy variable for 2003 birth in a regression and examined their ORs. For a robustness check on our measure of cumulative experience, we also ran the regressions with cumulative experience starting a year earlier (2001). Results remained the same. We assessed support for Hypothesis 2 by adding interdisciplinary collaboration and an interaction term consisting of DLA and interdisciplinary collaboration to our regression model and then examining the OR for the interaction term.

Concern about response bias related to workgroup performance was reduced by a two-sample t-test that showed no difference in 2001 SMR between the 23 NICUs in our sample and the 21 nonsampled NICUs. The average mortality for our sample was 15.4% in 2001, 14.6% in 2003, and 15.7% in 2004. t-Tests for differences in mortality between 2001 versus 2003, 2001 versus 2004, and 2003 versus 2004 were not significant. We also checked for differences between the characteristics (e.g., birth weight) for infants treated in 2001, 2003, and 2004 that might be associated with differences in mortality rates. t-Tests showed no significant difference in patient characteristics between 2001 and 2004 patients. Four infant characteristics were significantly different between the 2001 and 2003 infants (Pr\(T > t\) ≤ 0.10), which could account for the slightly lower, although statistically insignificant, mortality in 2003. In 2003, there was a smaller percentage of male infants, a higher percentage of infants with Hispanic mothers, a lower percentage with mothers whose race was classified as “other,” and higher mean Apgar scores after one minute. We include these characteristics as control variables in our regression equations, so the differences should not affect our results. All analyses were performed using STATA® version 10.0.

5. Results

5.1. Results of Descriptive and Correlation Analyses

Table 1 reports descriptive statistics for infant-level variables. The mortality rate for infants born in 2003 and 2004 in our sample was 15%. The average infant was almost 26 weeks’ gestational age and had a one-minute birth weight of 2596 g.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean or %</th>
<th>S.D.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mortality (2003 and 2004)</td>
<td>0.15</td>
<td>0.36</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1a. Mortality for 2003 infants only ((n = 1,998))</td>
<td>0.15</td>
<td>0.35</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1b. Mortality for 2004 infants only ((n = 1,061))</td>
<td>0.16</td>
<td>0.36</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2. Gestational age (weeks)</td>
<td>25.96</td>
<td>2.00</td>
<td>21</td>
<td>35</td>
</tr>
<tr>
<td>3. Birth weight (grams)</td>
<td>769.65</td>
<td>146.86</td>
<td>415</td>
<td>1,000</td>
</tr>
<tr>
<td>4. Small for gestational age</td>
<td>0.16</td>
<td>0.37</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5. Apgar score (1 minute)</td>
<td>4.74</td>
<td>2.31</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>6. Gender ((1 = \text{male}))</td>
<td>0.50</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7. Mother’s race—White</td>
<td>0.60</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8. Mother’s race—Black</td>
<td>0.23</td>
<td>0.42</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9. Mother’s race—Hispanic</td>
<td>0.10</td>
<td>0.30</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10. Mother’s race—Asian</td>
<td>0.05</td>
<td>0.21</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11. Mother’s race—Other</td>
<td>0.03</td>
<td>0.16</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12. Vaginal delivery</td>
<td>0.30</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13. Member of multiple birth</td>
<td>0.27</td>
<td>0.45</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14. Born at another hospital</td>
<td>0.26</td>
<td>0.44</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. There is no significant difference in mortality between 2003 and 2004 \((p\text{-values} > 0.10)\).
5.2. Results of Hypotheses Tests

The results of our analyses are shown in Table 3. Model 1, our first test of Hypothesis 1 across all infants in our sample, provides no support for our hypothesis; there was no association between the use of DLA and mortality, after accounting for cumulative experience, which also had no effect on mortality. Model 2 shows the results from our second test of Hypothesis 1, in which we examined the effect of using DLA by year. In this model, a one standard deviation increase in DLA was associated with an 18% increase in the odds of mortality in 2003 (OR = 1.18, p = 0.02) and an 18% decrease in 2004 (OR = 0.82, p = 0.02). Cumulative experience was not significant for 2003 infants but was for 2004 infants, with a one standard deviation increase in volume (13.2 ELBW infants per year) equal to a 20% reduction in the odds of mortality (OR = 0.80, p = 0.02). A χ² test comparing the 2004 ORs was not significant (χ² = 0.04, p = 0.84), suggesting that the mortality benefit of using DLA versus cumulative experience in 2004 was equal. These results provide partial support for Hypothesis 1.

5.3. A Reconceptualization: Examining a Mediating Rather than Moderating Effect

In making our moderation prediction, we extrapolated from research on interdisciplinary collaboration because there was limited research that discussed DLA and collaboration. Given our results, we speculated—in a modification of Hypothesis 2—that interdisciplinary collaboration mediates, rather than moderates, the positive relationship between the use of DLA and workgroup performance in the long term. In other words, DLA might foster interdisciplinary collaboration and workgroup performance in turn. Collaboration has been found to predict workgroup performance (Baggs et al. 1999). Alternatively, DLA could mediate between interdisciplinary collaboration and workgroup performance because interdisciplinary collaboration might foster the use of DLA, much the way psychological safety does (Tucker et al. 2007). Workgroups in which individuals collaborate may be more likely to use DLA. We tested these alternative hypotheses by assessing whether the
### Table 3 Results of Hypothesis Testing Using Logistic Regression Analyses Clustered by NICU ($N = 2,159$ Infants in 23 NICUs)

<table>
<thead>
<tr>
<th>Hypothesis:</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main effects only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variables</td>
<td>OR</td>
<td>RSE</td>
<td>OR</td>
</tr>
<tr>
<td>Independent variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative experience</td>
<td>0.93</td>
<td>0.06</td>
<td>0.94</td>
</tr>
<tr>
<td>Cumulative experience for 2003 infants</td>
<td>0.97</td>
<td>0.10</td>
<td>0.97</td>
</tr>
<tr>
<td>Cumulative experience for 2004 infants</td>
<td>0.80$^*$</td>
<td>0.07</td>
<td>0.80$^*$</td>
</tr>
<tr>
<td>Deliberate learning activities</td>
<td>1.01</td>
<td>0.07</td>
<td>1.14</td>
</tr>
<tr>
<td>Deliberate learning activities for 2003 infants</td>
<td>1.18$^*$</td>
<td>0.08</td>
<td>1.18$^*$</td>
</tr>
<tr>
<td>Deliberate learning activities for 2004 infants</td>
<td>0.82$^*$</td>
<td>0.07</td>
<td>0.82$^*$</td>
</tr>
<tr>
<td>Year of infant birth (1 = infant born in 2003)</td>
<td>0.84</td>
<td>0.11</td>
<td>0.84</td>
</tr>
<tr>
<td>Moderating variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interdisciplinary collaboration</td>
<td>0.76$^{**}$</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Deliberate learning activities x Interdisciplinary collaboration</td>
<td>1.00</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Control variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban location</td>
<td>1.19</td>
<td>0.19</td>
<td>1.33$^*$</td>
</tr>
<tr>
<td>Teaching hospital</td>
<td>0.66$^*$</td>
<td>0.12</td>
<td>0.66$^*$</td>
</tr>
<tr>
<td>Improvement team size</td>
<td>0.84$^*$</td>
<td>0.06</td>
<td>0.85$^*$</td>
</tr>
<tr>
<td>Staff tenure</td>
<td>1.20$^*$</td>
<td>0.10</td>
<td>1.18$^*$</td>
</tr>
<tr>
<td>Prior performance (2001 SMR)</td>
<td>1.29$^*$</td>
<td>0.13</td>
<td>1.28$^*$</td>
</tr>
<tr>
<td>Pseudo $R$-squared (%)</td>
<td>0.149</td>
<td>0.154</td>
<td>0.155</td>
</tr>
<tr>
<td>Obs. correctly classified (%)</td>
<td>85.18</td>
<td>85.27</td>
<td>85.36</td>
</tr>
</tbody>
</table>

Notes. An odds ratio (OR) less than 1 indicates lower mortality; conversely, an OR greater than 1 indicates higher mortality. Robust standard errors (RSEs) have been corrected for clustering of infants within NICUs. All models include infant-level control variables used for risk adjustment (Rogowski et al. 2004) and constants that are not reported for simplicity. Continuous variables have been standardized to facilitate comparison of effect across variables. Staff tenure was measured as follows: 1 = <1 year, 2 = 1 year to 2 years, 3 = 2–5 years, 4 = 5–10 years, 5 = >5 years.

$
^{*}p \leq 0.10; ^{**}p \leq 0.05; ^{***}p \leq 0.01.
$

Results from our mediation analysis are shown in Table 4. We first tested whether interdisciplinary collaboration mediated the relationship between the use of DLA and workgroup performance. Model 1 (identical to Model 2 in Table 3, repeated for clarity) shows that the first criterion for mediation was satisfied: a positive relationship existed between use of DLA and patient mortality for 2004 infants. Model 2 shows that the second criterion was satisfied: use of DLA was positively associated with interdisciplinary collaboration. Because collaboration was a continuous variable, Model 2 shows $\beta$-values from an ordinary least squares (OLS) regression analysis. Model 3 shows that the final criterion for mediation was satisfied. In this model, which included both use of DLA and interdisciplinary collaboration, the significant positive relationship between use of DLA and patient mortality for 2004 infants disappeared, whereas the significant relationship between interdisciplinary collaboration and patient mortality remained (OR = 0.82, $p = 0.05$). With 44% of the effect mediated, a Sobel test using the infants treated in 2004 was also significant ($p = 0.003$). Thus our results support the reconceptualization of interdisciplinary collaboration as a mediator, rather than moderator, of the relationship between use of DLA and workgroup performance.

Our results did not support the alternative hypothesis that DLA mediated the relationship between interdisciplinary collaboration and workgroup performance. Model 4 shows that the first condition for mediation was satisfied: collaboration was positively associated with patient mortality (OR = 0.79, $p = 0.006$). However, Model 5 shows that the second condition was not satisfied: collaboration did not predict the use of DLA ($\beta = -0.01, p = 0.92$). A Sobel test was also not significant ($p = 0.38$).

### 6. Discussion and Conclusions

Our results contribute to management theory, research, and practice in three ways. First, we extend theory and
Table 4 Results of Mediation Regression Analyses Clustered by NICU

<table>
<thead>
<tr>
<th>Regression analysis used</th>
<th>Dependent variable</th>
<th>Criterion for mediation being tested</th>
<th>Coefficients reported</th>
<th>Mediation reconceptualization 1</th>
<th>Mediation reconceptualization 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Use of DLA → Collaboration → Mortality</td>
<td>Use of DLA → Collaboration → Mortality</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Logistic</td>
<td>Mortality</td>
<td>Criterion 1</td>
<td>OR, RSE, β, RSE</td>
<td>0.97</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Criterion 2</td>
<td></td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Criterion 3</td>
<td></td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Logistic</td>
<td>Collaboration</td>
<td>Criterion 1</td>
<td>OR, RSE, β, RSE</td>
<td>0.21**</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Criterion 2</td>
<td></td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Criterion 3</td>
<td></td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Logistic</td>
<td>Mortality</td>
<td>Criterion 1</td>
<td>OR, RSE, β, RSE</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Criterion 2</td>
<td></td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Criterion 3</td>
<td></td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Logistic</td>
<td>Interdisciplinary collaboration</td>
<td>Criterion 1</td>
<td>OR, RSE, β, RSE</td>
<td>0.04</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Criterion 2</td>
<td></td>
<td>0.11</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Criterion 3</td>
<td></td>
<td>0.09</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Notes. β is the standardized OLS regression coefficient. RSEs have been corrected for clustering of infants within NICUs. Models 1, 3, and 4 include infant-level control variables used for risk adjustment (Rogowski et al. 2004) and constants that are not reported for simplicity. Models 2 and 5 contain control variables for profession and gender. The number of observations differs for these models because the data are individual NICU staff members (n = 1,081) and NICU improvement team members (n = 244), respectively. Continuous variables have been standardized to facilitate comparison of effect across variables. Staff tenure was measured as follows: 1 = <1 year, 2 = 1 year to 2 years, 3 = 2–5 years, 4 = 5–10 years, 5 = >5 years.

*p ≤ 0.10; *p ≤ 0.05; **p ≤ 0.01.

Research on organizational learning in dynamic service settings by examining the contribution of induced learning to performance improvement in this prevalent, yet understudied setting. Second, we provide evidence on the role of a key workgroup characteristic, interdisciplinary collaboration, in the relationship between induced learning via DLA and workgroup performance. Third, our findings have practical significance because they reveal the effect over time of a strategy for improving quality of care in health-care organizations, which have a well-documented need for such improvement. We discuss our results and their implications with respect to each contribution.

6.1. Extending Theory and Research on Organizational Learning in Dynamic Service Settings

Despite the prevalence and centrality of dynamic service settings in our economy and society, little research has examined organizational learning in these settings. The studies that have been conducted in this setting—some by management scholars and some by health services researchers—have largely investigated autonomous learning via cumulative experience (e.g., patient volume) as a strategy for performance improvement. Whether induced learning through the use of DLA might provide additional performance benefits in this setting has been neglected, with the exception of qualitative work by Pisano et al. (2001). We examined the use of DLA in the dynamic setting of hospital units and found that they provide benefits after a two-year period, even after accounting for cumulative experience. Specifically, we found that after two years, a one standard deviation increase in the use of DLA improved workgroup performance by 18%. In the shorter term, using DLA had a negative effect on performance. Our results reinforce the importance of including induced learning alongside autonomous learning in organizational learning models of dynamic service settings and considering how their effects change over time.
Our finding of a worse-before-better effect suggests that using DLA initially presents a challenge for many workgroups. Research by Keating et al. (1999) supports this hypothesis. In their study of DuPont and other large manufacturers, Keating et al. (1999) observed that firms’ improvement programs often led to declining business performance in the short term, a phenomenon they termed the “improvement paradox.” They theorized that this phenomenon occurs because of a trade-off between improvement and production. Without additional resources, workgroup investment in improvement creates less capacity to fulfill current production demands; thus, performance for current tasks declines until workgroups are able to compensate. A similar challenge likely exists for workgroups using DLA. Performance falls until workgroups implement the improved practices they learned. Therefore, an implication of our study is that workgroups must be vigilant during initial phases to minimize adverse effects on performance.

Furthermore, our study suggests that induced learning should be considered nearly equal to autonomous learning in its long-term effect on workgroup performance in our setting. For the 2004 infants, we found no significant difference between the level of improved workgroup performance associated with a one standard deviation increase in the use of DLA versus cumulative experience (i.e., 18% compared to 20% reduction in patient mortality, respectively). The finding of a similar effect size is consistent with findings from research conducted on the learning of new production processes in manufacturing settings. Hatch and Mowery (1998) found that the coefficients on cumulative volume and investment in engineering analysis in their model of manufacturing yield rates were similar in magnitude. Likewise, in a study of product defect rates, Ittner et al. (2001) found a statistically insignificant difference in the effect of induced learning as measured by past expenditures on conformance (13%) and the effect of autonomous learning as measured by past cumulative volume (7%). Thus although there are a number of characteristics that differentiate manufacturing from service settings, there are ways in which the settings are alike. Our study shows that there is similarity in the longer-term impact of autonomous and induced learning in each setting, at least with respect to learning new practices. The similar effectiveness of each process means managers have potentially substitutable learning tools at their disposal, which is beneficial given the “paucity of experience problem” in dynamic settings identified by Levitt and March (1988, p. 333). Our findings suggest that managers may circumvent this problem without sacrificing long-term performance by facilitating induced learning.

6.2. The Role of Interdisciplinary Collaboration in Induced Learning Effectiveness

Our results suggest that interdisciplinary collaboration within the workgroup is a key variable in understand-
6.3. An Effective Strategy for Improving Quality of Health Care

The well-documented quality problems in health-care organizations have created an imperative to identify effective strategies for improving quality of care. Although the strategy of autonomous learning has offered some benefits, as evidenced by studies linking cumulative volume to patient mortality, it has not uniformly enhanced the level of performance for health-care organizations (Pisano et al. 2001, Reagans et al. 2005). As a result, health-care professionals and organizations have sought information on alternative strategies they may use to enhance their performance. Our results suggest that they may improve quality of care in the long term (though not in the short term) by using the seven DLAs we examined. Prior research showed that hospital staff perceived that the use of this bundle of activities contributed to the successful implementation of new practices in their units (Tucker et al. 2007). That research, however, did not examine the relationship between perception and reality. Our results extend the prior work by providing empirical evidence of the long-term effectiveness of this bundle of activities with respect to a key objective metric of service quality in health care: patient mortality. They also indicate, however, that staff must be hypervigilant during the initial deliberate learning period to ensure that these activities do not result in negative outcomes in the short term. Instituting the principles for high reliability in high-risk settings (e.g., sensitive operations; see Weick et al. 1999 for a discussion) may be helpful in this regard.

6.4. Limitations and Suggested Future Research

Like any study, this one has limitations. First, it only included 23 of the 40 NICUs in the collaborative eligible to participate in the study. The results of our independent samples t-test showing no difference in structural and clinical measures between the 23 participants and the 21 nonparticipants (including the four pilot sites) mitigate concerns about selection bias. However, we cannot eliminate the possibility that study participants differ in ways that prevent the generalizability of our findings to NICUs not in the collaborative. Second, because we studied only NICUs, a specialized hospital unit, the generalizability of our findings to other health-care or non-health-care settings with different characteristics is unknown. Differences in the level of professional education required by the workgroup, criticalness of the workgroups’ actions (not every setting involves life or death decisions), degree of status differences within the workgroup, and length of relationship with the customer may make our findings more or less relevant. Finally, we must note that our data on the use of DLA and interdisciplinary collaboration were cross-sectional. Thus although we controlled for prior performance when examining recent performance, which is a strength of this study (West et al. 2006), we cannot make causal inference. We do not know whether the use of DLA truly enabled better interdisciplinary collaboration. An omitted variable, such as financial resources, could have triggered the use of DLA, collaboration, and better performance (Huesl 1995, Klein et al. 2001). Also, there may be feedback loops between our variables that we could not detect (e.g., DLA leading to collaboration, which in turn motivates more DLA).

Future research should pursue longitudinal measures to shed insight on three unanswered questions. First, future research is needed to yield deeper understanding of the temporal aspect of induced learning. Our study found that performance first worsened and then improved after the use of DLA. Future studies should track both DLA and performance for at least three years to determine whether the positive impact of these activities continues and for how long. Second, future research with a sample size of at least 40 groups is needed to explore the possibility that the differential effect of DLA and cumulative experience by year masks differences in effectiveness of each approach for different groups. For example, poorly performing workgroups may benefit from autonomous learning, whereas higher-performing groups may benefit from induced learning. Determining whether there is a nuanced approach to performance improvement would benefit collaboratives, which to our knowledge do not tailor their program of improvement activities based on each workgroup’s current performance. Third, gathering measures of interdisciplinary collaboration at multiple points in time would answer questions about the causality between DLA and interdisciplinary collaboration. We suspect that the relationship is mutually reinforcing, creating a positive dynamic cycle.

We hope that future research will also shed light on which mechanism(s) underlying collaboration’s effectiveness—better decision making, better coordination, or better error detection and recovery—explain interdisciplinary collaboration’s relationship to DLA. Such research would require detailed observations to illuminate differences in how teams with varying levels of DLA work together. For example, a study in the health-care setting might examine different teams performing direct patient care (e.g., surgery) to see how the nature of interdisciplinary collaboration changes following use of DLA. Future research should also evaluate the comparative effectiveness of the use of DLA versus other actions managers might take to promote collaboration, such as superordinate goals (Pinto et al. 1993) and incentives (Siemsen et al. 2007). Studies are also needed to identify other mechanisms through which DLA impact workgroup performance.

Future research could also examine potential moderators. Although our study failed to show that interdisciplinary collaboration moderates the relationship between DLA and workgroup performance, this relationship may be moderated by other factors. There are probably also moderators of the relationship between interdisciplinary
collaboration and performance, such as the level of uncertainty in work. For example, it is possible that interdisciplinary collaboration has a larger impact in crisis situations (higher level of uncertainty) than in routine work (lower level of uncertainty). Another possible moderator is the rate of knowledge change. It is likely that DLA are more important when the focal knowledge base changes rapidly, as it does in neonatal intensive care.

Future research also should consider other measures of performance. Although patient mortality is a key performance measure for health-care organizations, it is an extremely blunt measure of quality. More fine-grained measures (e.g., incidence of adverse events) might further illuminate differences in quality of care derived from workgroup attributes such as use of DLA and interdisciplinary collaboration. One drawback to such measures is the difficulty of gathering uniform data across organizations. However, the movement toward national standardized performance reporting will improve access to such data. Future research should capitalize on these data to investigate the effects of DLA and interdisciplinary collaboration on less blunt outcome measures. Studies using other measures of service performance (e.g., patient satisfaction and length of stay) could offer further insight on organizational learning in health care.

6.5. Conclusion
In sum, this study contributes to the theoretical and empirical literature on organizational learning by demonstrating the importance of DLA for better workgroup performance in dynamic service settings such as hospitals. Research in other settings had suggested that induced learning has performance benefits, but relatively little empirical research has been conducted to explore this claim in dynamic service settings. Our results suggest that induced learning via the use of DLA does improve performance after two years. Moreover, our results provide insight into why the use of DLA has this effect on performance. Our data suggest that the use of DLA cultivates interdisciplinary collaboration among staff that must work together to perform well. This finding is particularly useful for managers who have struggled to identify concrete means for increasing interdisciplinary collaboration. Our results revealed seven specific learning activities that form a cohesive bundle for deliberate learning and interdisciplinary collaboration in turn. We hope that this knowledge aids organizations with the challenging task of performance improvement. We also hope that this research provides insight for future studies aimed at developing a better understanding of deliberate learning and other strategies for performance improvement in dynamic settings.

Acknowledgments
This research would not have been possible without the collaboration of the Vermont Oxford Network (VON), which granted access to its patient outcome data for the NICUs in this study. In particular, the authors are deeply appreciative of the unfailing support provided by Jeffrey Horbar, Chief Executive and Scientific Officer of the VON. The authors benefited greatly from consultations with him about the NICU environment and this research. The authors are also grateful to Joseph Carpenter, VON Operations Director and Statistician, for providing the outcomes data set. Amy Edmondson and Richard Bohmer of Harvard Business School provided valuable advice and assistance throughout the study. The authors also thank Senior Editor Ray Reagans, two anonymous reviewers, Judith Gedney Baggs, Jane Banaszak-Holl, Anat Drach-Zahavy, Joy Field, Michael Lapré, Enno Siemsen, Michael Toffel, Timothy Vogus, and Michelle Williams for their insightful comments on earlier versions of this work. This project benefited from the research assistance of Laura Feldman and the statistical advice of William Simpson. The Institutional Review Board at Harvard University approved this research and the Harvard Business School Division of Research, Wharton’s Fishman-Davidson Center for Service and Operations Management, and a Sloan Industry Studies Fellowship provided financial support. This paper was selected for inclusion in the Best Paper Proceedings of the 2010 Academy of Management Meeting. The authors gratefully acknowledge the NICUs in the following hospitals for their participation in this study: Advocate Lutheran General Children’s Hospital; Children’s Hospital of Dartmouth, Children’s Hospital of Orange County; Children’s Hospital of Southwest Florida, Children’s Hospital Medical Center of Akron, Janet Weis Children’s Hospital at Geisinger Medical Center, Mission Children’s Hospital, Parkview Hospital, Providence St. Vincent Medical Center, Rockford Health System, St. John Mercy Medical Center, The Children’s Hospital at Providence, Woman’s Hospital, and Yakima Valley Memorial Hospital, as well as those in hospitals that preferred to remain anonymous.

Appendix. Survey Items
For each survey item, respondents reported their level of agreement on a scale ranging from 1 = strongly disagree to 7 = strongly agree.

A.1. Interdisciplinary Collaboration
1. Nurses and physicians work well together.
2. Communication between nurses and physicians is open and positive.
3. When there is a disagreement between nurses and physicians, all points of view will be carefully considered in arriving at the best solutions to the problem.
4. Overall, our unit functions very well together as a team.
5. The input of respiratory therapists and/or other ancillary staff is regularly sought when developing treatment plans.
6. There is a good understanding of each other’s job responsibilities among all those involved in the care of patients in the NICU.

A.2. Psychological Safety
1. People in this unit are comfortable checking with each other if they have questions about the right way to do something.
2. The people in our unit value others’ unique skills and talents.
3. Members of this NICU are able to bring up problems and tough issues.
Endnotes
1 We elected to use nonphysicians’ responses because prior work showed that physicians’ estimates of work climate differ significantly from that of other disciplinary groups. Moreover, unlike nurses’ estimates, they do not predict patient outcomes (Baggs et al. 1999).
2 We thank our reviewers for suggesting this alternative conceptualization.

References


---

Igrid M. Nembhard is an assistant professor of public health and management at Yale University’s School of Medicine (Public Health) and School of Management. She received her Ph.D. in health policy and management, with a concentration in organizational behavior, from Harvard University. Her research focuses on organizational learning and the implementation of new practices in health-care organizations.

Anita L. Tucker is an associate professor in technology and operations management at the Harvard Business School. She received her doctorate in business administration from the Harvard Business School. Her research addresses organizational learning and improvement on the frontlines of complex service organizations.