Accepted Manuscript

DOI: https://doi.org/10.1111/poms.12874

Citation: Avgerinos, E., & Gokpinar, B. (2018). Task variety in professional service work: When it helps and when it hurts. Production and Operations Management, 27(7), 1368-1389.

Task Variety in Professional Service Work: When It Helps and When It Hurts

In a wide range of professional service firms, individuals perform a variety of tasks which are highly cognitive and knowledge intensive yet repetitive in nature, providing significant opportunities for learning. In addition, individuals in such environments tend to enjoy considerable discretion in managing when and how they perform their tasks. In light of these observations, we investigate task allocation and timing strategies that may enhance or inhibit learning and productivity for professional service workers. Specifically, we focus on the role of task variety. We use a detailed dataset of 3,275 coronary artery bypass surgeries in a private European hospital over seven years to examine the effect of concurrent and non-concurrent exposure to task variety on learning and productivity on a focal task. We find that while concurrent exposure to variety has a positive impact on focal productivity, non-concurrent exposure to variety has a negative impact on it. Our results also suggest that short term exposure to variety amplifies these relationships.

Key words: professional service work, task variety, learning, productivity, task timing.

1. Introduction

Professional service firms globally generate annual sales over $3 trillion and represent 7-8% of total service sector revenue in advanced economies (McKinsey 2012). This percentage is even higher in service-based economies such as Britain, where 15% of GDP and 14% of employment comes from professional services firms (PwC 2012). Following Von Nordenflycht's (2010) characterization, the professional service industry broadly includes accounting, advertising and marketing, management consulting, architecture, legal services, scientific research services, and physician practices. While these firms have distinct characteristics at a high level, including knowledge intensity, low capital intensity, and a professionalized workforce (Von Nordenflycht 2010), they also demonstrate several key features from an operations standpoint in terms of the way their employees perform their work.

First, the majority of the work performed by professional service workers is quite repetitive in nature. For example, management consultants follow similar steps in their engagement with clients from initiation to contracting and final deliverables; legal professionals draft legal documents by engaging in a similar set of activities; and surgeons perform in the operating room by following a certain set of procedures. While most activities and tasks may be quite similar from one job to another, one still observes high cognitive activity among workers, presumably due to the variation in work content across tasks (e.g., differences between consulting projects, between surgeries, or legal cases). Hopp et al. (2009) consider this key observation in their classification of white-collar work: they call these (e.g., consulting, legal services) intellectual and routine work. As a result of the repetitive nature of the work and the significant opportunities for learning that these settings offer,

individuals inevitably transfer their past experience and knowledge when they work on subsequent tasks (Tversky 1977, Gick and Holyoak 1987, Zollo and Reuer 2010).

A second important feature of professional service work is that individuals tend to have a relatively high degree of discretion in managing when and how they perform their tasks (Hopp et al. 2007). Compared with workers in other professions, they enjoy more control and flexibility over decisions regarding task sequences, including whether to perform certain tasks concurrently or individually and in smaller pieces or larger chunks.

Considering these two key features of professional service work—first, its repetitive nature and many learning opportunities across tasks and, second, workers' potential discretion in managing when and how to perform various tasks—an important operational question is the following: How should professional service workers perform their various tasks to achieve greater learning and productivity over time? More specifically, in performing the variety of tasks that a professional service worker is supposed to carry out, are there certain task timing configurations that enhance or inhibit learning and productivity?

Our study seeks to address these questions by focusing on the way various tasks are performed. We distinguish between concurrent and non-concurrent exposure to variety and examine productivity implications of these two approaches to organizing work. *Concurrent variety* refers to performing another task concurrently with the focal one, whereas *non-concurrent variety* refers to performing another task independently (i.e., at a different time) from the focal one. Because many professional service workers inevitably perform a variety of tasks, we seek to understand when and how exposure to variety helps, and when and how it hurts individuals' focal productivity.

Exposure to variety through successful knowledge transfer to the focal task may have a positive impact on performance (Schilling et al. 2003, Boh et al. 2007, Staats and Gino 2012), but too much exposure to variety may be detrimental to productivity (Narayanan et al. 2009). Also, task variety could be confusing for individuals (Allport et al. 1994) and therefore decrease their subsequent focal productivity due to switching costs and warm-up periods (Cellier and Eyrolle 1992, Monsell 2003). Like the studies that have produced these findings, our study explores the productivity implications of exposure to task variety, but with an important distinction. We propose that the influence of task variety on productivity critically depends on the way (i.e., when) other tasks are performed in relation to the focal task. We suggest that knowledge transfer and learning mechanisms are quite different when other tasks are performed concurrently vs. non-concurrently with the focal task, which leads to differentiated and, in fact, contrasting effects on productivity.

Specifically, we develop and test four hypotheses regarding professional service workers' concurrent and non-concurrent exposure to variety by examining their effect on productivity in subsequent focal tasks. We find that concurrent exposure to variety enhances productivity, whereas non-concurrent exposure to variety is detrimental to productivity. Concurrent variety is beneficial for productivity because it is highly conducive to learning: it enables 'implicit learning' (Reber 1989,

Wulf and Schmidt 1997) and facilitates cognitive skill acquisition through the discrimination process (Anderson 1982), all of which leads to better comprehension of the focal task. Non-concurrent variety, on the other hand, results in reduced productivity through negative transfer effects due to the difference between surface and structural similarities (Gick and Holyoak 1987) and priming (Allport et al. 1994). In addition, our results suggest that short-term exposure to variety amplifies the influence of the long-term ones on subsequent focal productivity. That is, recent (short-term) concurrent variety increases the positive influence of concurrent variety on focal productivity, whereas recent non-concurrent exposure to variety strengthens the negative impact of non-concurrent variety on focal productivity.

As we test our hypotheses and explore how professional service workers can achieve higher productivity through better allocation of a variety of tasks they perform, we should emphasize an important empirical consideration. Because many service workers have significant discretion over when and how to perform their tasks, there may be inherent endogeneity in many professional service settings, which could pose problems in empirical identification. In order to test the hypotheses developed in the present paper, ideally one needs a professional service setting where decisions about which tasks to perform, when to perform different tasks, and whether to perform them concurrently or non-concurrently are not up to individuals' discretion, but instead are determined exogenously. We test our hypotheses using a detailed dataset of 3,275 coronary artery bypass graft (CABG) operations from the cardiac unit of a private European hospital over seven years. Because a patient's need is the primary driver of the type, nature, and timing of an operation and these are not up to the discretion of the surgeon, we believe our setting is ideal in which to investigate the role of task variety on productivity.

Our paper offers a number of significant contributions to the service operations management literature. First, by focusing on an under-studied service sector, namely operations of professional services firms (Roth and Menor 2003, Lewis and Brown 2012), our study examines how different task allocation and composition strategies may influence learning and subsequent focal productivity of professional service workers. This, we believe, is also a step towards answering Argote et al. (2003)'s call for more research to identify "mechanisms and conditions under which experience is beneficial (or harmful) for learning outcomes" and a step towards answering their question about whether different types of experience may provide better understanding of the task (p. 579). Second, our paper contributes to the growing literature on task variety and productivity by introducing a new dimension of variety which has not been considered before—that is, concurrent and non-concurrent exposure. In a similar study, Staats and Gino (2012) focused on a single current task, and studied when different tasks take place (on the same day or in the past) with respect to that current task, and examine how these affect current task performance. We, however, concentrate on a common focal task being performed over time (i.e., CABG), and study *how* exposure to variety takes place with respect to past focal tasks—in conjunction with it (concurrently) or independently from it (non-concurrently)— and

examine their subsequent performance implications. Third, we make an important distinction between short-term and long-term learning dynamics. We introduce short-term exposure to variety as a factor that moderates how long-term exposure to variety (both concurrent and non-concurrent) affects individuals' focal productivity.

In the next section, we describe our setting before proceeding to motivate and develop our hypotheses.

2. Setting

Our setting is the cardiac unit of a private hospital in Europe, and our dataset consists of 3,275 coronary artery bypass graft (CABG) surgeries that were conducted in the hospital over a period of seven years and three months. In addition to CABG surgeries, which we identify as the focal task (see Section 4 for a discussion), we also have information regarding all other types of cardiac surgeries that were conducted in the hospital during the same time period, which allows us to observe surgeons' exposure to other types of tasks (i.e., task variety).

Our setting is a very suitable context in which to investigate the effects of exposure to variety on individual learning and productivity because surgeons perform CABG surgeries as well as a variety of other types of cardiac surgery and can therefore potentially transfer knowledge from one task to another. In addition, learning is an integral part of hospital operations (Tucker et al. 2007) and surgeons' practices (KC and Staats 2012). Furthermore, CABG surgeries are highly critical and complex yet quite common and frequent tasks for surgeons (Pisano et al. 2001, Clark and Huckman 2012), making them an ideal setting in which to study how exposure to task variety may influence learning and the resulting productivity of professional service workers. Finally, because our dataset covers a span of more than seven years, we are able to examine both long-term and short-term effects of task variety on learning and subsequent focal productivity.

As mentioned previously, one major advantage of our setting is that the nature, type, and time of the tasks (surgeries) are not endogenously determined by the worker but are driven by outside factors (patients' needs). Consequently, a surgeon may perform a single CABG surgery if that is all the patient needs; she may perform a single valve replacement if the patient requires only that, or she can perform multiple surgeries during the same operation (valve replacement and CABG) if the patient needs both. Likewise, surgeons' assignments to tasks and similarity of their skills sets are other important considerations to avoid endogeneity, which we investigate in detail in the robustness checks section.

In this setting, we identify CABG as the focal surgery and examine the impact of surgeons' exposure to other types of surgeries on subsequent focal task (CABG) productivity. Because surgeons perform other types of surgeries, too, both concurrently (e.g., performing valve replacement and CABG together) and non-concurrently (e.g., performing a single valve replacement) depending on

medical requirements, we are able to observe knowledge transfer and learning between both concurrent and non-concurrent tasks.

3. Literature Review and Theory Development

3.1 Exposure to Variety, Learning, and Productivity

Task variation can be established by performing either related tasks (related variation) or unrelated tasks (unrelated variation) (see Schilling et al. 2003 for a discussion of this distinction).[1] Organizational research that examines the performance implications of exposure to related task variety at the individual, team, or organizational level has found mixed results. Clark and Huckman (2012), for example, found no evidence that related activities generated positive spillovers on the focal activity within operating units in different hospitals. Narayanan et al. (2009) found that variety had an inverted U-shaped relationship with individual productivity in an offshore software support services company. On the firm level, Haleblian and Finkelstein (1999) suggested that prior related experience has a U-shaped relationship with the focal experience in the context of acquisitions. Finally, KC and Staats (2012) found that focal subtask variety has an inverted U-shaped relationship with performance, whereas related subtask variety has a U-shaped relationship with the performance of cardiac surgeons. Finally, Ibanez et al. (2016) use data from physicians reading of radiology studies and show that when individuals exercise discretion and deviate from the assigned order, they tend to experience slower reading times.

A significant body of literature argues that task variety enhances learning through successful transfer of knowledge between the focal and related activity. The more similar these activities are, the higher the probability of a successful knowledge transfer and application (Tversky 1977, Zollo and Reuer 2010). In addition, task variety can increase workers' commitment and motivation (Hackman and Oldham 1976, Langer 1989), resulting in improved productivity. Boh et al. (2007) use data from the software industry to show that, on a team and organizational level, exposure to related systems is more beneficial for performance than specialization. Schilling et al. (2003) show that related variety can improve the learning rate of students playing different versions of a game. Staats and Gino (2012) use data from a Japanese bank to show that variety promotes workers' productivity in the long run.

On the other hand, researchers have also argued that task variety could be distracting and eventually detrimental to individual learning and productivity (Allport et al. 1994, Monsell 2003). Prior learning in one task may have a negative effect in performing another related task due to "negative transfer effect" which is well established in cognitive psychology literature at the individual level (Gick and Holyoak 1987, Zollo and Reuer 2010). Holyoak (1985) shows that when there is a difference between an individual's perception of similarity and the actual similarity of the tasks, a

[1] Our study only considers related variation, since most professional service workers perform highly related tasks as part of their jobs (e.g., preparing various legal cases or performing related surgeries). Consequently, throughout this paper, when we say "other task", we refer to a task that is different from but related to the focal one.

negative transfer effect of knowledge can appear through invalid generalizations. Similarly, Cohen and Bacdayan (1994) found that individuals who played a card game had performance difficulties when the rules were slightly changed.

3.2 Concurrent and Non-Concurrent Exposure to Variety

Our paper contributes to this debate by suggesting and testing a new mechanism by which task variety may influence individual learning and focal productivity. We reconcile the two views outlined above and suggest that exposure to variety independently (non-concurrently) versus concurrently determines the direction and level of successful knowledge transfer from the related task to the focal task. That is, non-concurrent task variety could result in negative knowledge transfer and hence lead to lower productivity the next time the focal task is performed. On the other hand, concurrent variety could enhance knowledge transfer from related tasks to the focal task and hence improve individual productivity for subsequent focal task.[2] For example, during an operation, a surgeon might perform only a valve replacement or might perform a valve replacement combined with a CABG. In the former operation, she will be exposed to variety (i.e., valve replacement) non-concurrently with the focal task (i.e., CABG), whereas in the latter case her exposure to variety will occur concurrently with the focal task.

3.3 Non-Concurrent Exposure to Variety

Previous studies on the positive effects of exposure to task variety on performance highlight two basic mechanisms. These include motivational benefits due to task variety (Herzberg 1968, Hackman and Oldham 1976) and positive knowledge transfer as a result of exposure to variety (Monsell 2003, Zollo and Reuer 2010). Cardiac surgery settings are high-pressure and dynamic environments (Edmondson 1999, Tucker and Edmondson 2003, Nembhard and Edmondson 2006, Wetzel et al. 2006), therefore, there is not much motivational gain to be realized as a result of exposure to variety (KC and Staats 2012). Also, because cardiac surgeries are highly complex tasks that demand a combination of cognitive and motor skills (Schaverian 2010), we do not expect to see straightforward positive knowledge transfer from task variety, especially when other tasks are performed non-concurrently. In fact, in the following lines we argue why we expect negative transfer effects from non-concurrent variety which will be detrimental for focal task productivity.

Negative transfer effects between tasks have been well established in the cognitive psychology literature (see Gick and Holyoak, 1987 for a review) in that prior learning and experience in one task may produce negative transfers and hence may be misleading in executing another task. In explaining

[2] Note that, the question we examine is fundamentally different from that of Staats and Gino (2012), who examine the performance implications of "same-day different task" and "all prior days' different tasks". By contrast, we build on the observation that knowledge workers can perform various tasks either separately (i.e., non-concurrently) or together (i.e., concurrently) and highlight how this differential way of getting exposure to different tasks influences subsequent productivity on a common focal task.

such negative transfer effects, researchers have highlighted a critical distinction between surface and structural similarities between tasks (Gick and Holyoak 1987, Novick 1988). Surface similarity (i.e., perceived similarity) is characterized by similarity of context (Craik and Tulving 1975), goals, and processes between two tasks (Bransford and Franks 1976, Tulving and Thomson 1973). Even though the specific elements involved in such tasks may resemble each other, underlying causal structure and ordering of elements could be quite different (Novick 1988). In contrast, structural similarities (i.e., objective structural similarity) refers to the deeper association between the elements of two tasks and their structural relatedness (Gick and Holyoak 1987) where one can derive causal inferences for the other task (Holyoak and Koh 1987). They key tenet of the distinction between surface and structural similarities is the emergence of negative transfer effects between tasks which have similar surface elements but different structural ones (Holyoak 1985). This suggests that experience in one task can be a liability for another task which is similar on surface but not-so-similar in the underlying structure. Based on this literature, we argue that while most cardiac operations are similar at the surface, there may be important structural differences between them which may lead to a negative transfer. Despite having seemingly many similarities in terms of the context, aim and process (i.e., surface similarity), each cardiac operation is carried out in distinct steps (Reznick and MacRae 2006). In fact, among various cardiac operations, medical literature has identified valve operations as being most similar to CABG in terms of the overlap in the surgical process and skills required (Ch'ng et al. 2015). Yet, in a recent large scale medical study involving 23 hospitals, 109 surgeons, and 20,619 patients over 10 years, Ch'ng et al. (2015) found (among others) no beneficial effect of surgeons' valve operation experience on their CABG performance. This was in contrast to their hypothesized positive effect. This indicates that, even having experience in an operation with very high structural similarity to CABG, that is experience in valve operation, was not found to be significantly associated with outcomes of CABG, which we suggest is due to negative transfer effects. If this is the case, then we expect to observe greater negative transfer effects from other types of cardiac operations which have much lower structural similarity to CABG. Such negative transfer from non-concurrent variety will result in decreased productivity (Novick 1988, Nokes 2009, Zollo and Reuer 2010) for CABG operations.

In addition to negative transfer effects, development of implicit memory (i.e., priming) due to non-concurrent variety can impair decision making in a subsequent focal task. Considering that medical researchers identify surgery operations to require twenty-five percent technical and seventy-five percent decision-making skills (Spencer 1978, Grierson et al. 2011), this could have important implications. Individuals evoke different sets of actions in response to a stimulus when performing another task (Allport et al. 1994, Allport and Wyllie 1999, Wyllie and Allport 2000, Waszak et al. 2003, Staats and KC 2012), and researchers have shown that unconscious response from past stimuli does occur among doctors and medical staff (Loewenstein and Lerner 2002, Bargh and Williams 2007). This unconscious response not only interferes with and lengthens the focal task (Allport et al.

1994), but also can lead to suboptimal decisions and action sets in the focal surgery (i.e., CABG). That is, priming due to non-concurrent variety may result in surgeons responding to different but related surgery types with a number of action sets, some of which may be inappropriate and detrimental for the focal task. Also, the number of action sets that surgeons automatically adopt when facing a related experience will increase over time, and this may in turn decrease surgeons' productivity when they perform the focal task. For these reasons, we predict that:

Hypothesis 1: *Non-concurrent exposure to task variety has a negative impact on subsequent focal task productivity.*

3.4 Concurrent Exposure to Variety

We next examine the productivity implications of performing other tasks in conjunction with the focal task. When another surgery is performed concurrently with the focal one, the dangers of invalid generalizations over time and negative learning transfer are significantly reduced. This is because the individual is also performing the focal surgery during the same operation, and this concurrent surgery will put the surgeons in a state of mindful activity and high alertness (Levinthal and Rerup 2006, KC 2014). This, in turn, will enable them to spot the nuances of and differences between the two types of surgery more easily, hence reducing the likelihood of invalid generalizations and negative learning transfer. In addition, the difficulty of performing different operations concurrently leads to enhanced information processing and decision-making ability, which promote efficient learning (Shea and Zimny 1983, Lee and Simon 2004). Indeed, recent medical research has observed that performing multiple surgeries during the same operation can promote more efficient learning for trainee surgeons (Bongers et al. 2015) and that surgeons performing multiple tasks at once are able to successfully reallocate their attention resources (Grierson et al. 2011). That is, performing another task concurrently with the focal one (e.g., performing both CABG and valve replacement concurrently) helps surgeons better comprehend, learn about, and identify the intricacies of the focal task (CABG) itself. We identify two mechanisms of this effect.

First, performing a focal task concurrently with other tasks provides a new context for the focal task. This variation in context will help surgeons develop "implicit learning" (Reber 1989, Wulf and Schmidt 1997). That is, without even realizing it, surgeons will develop critical but highly complex and abstract knowledge about the focal task and its associations (Maskarinec and Thompson 1976). Through this implicit learning process, surgeons will improve their understanding and performance of the focal task. Indeed, researchers have shown that implicit learning promotes neural efficiency (i.e., more expert-like mapping of neural resources in the completion of the task) in surgical training (Zhu et al. 2011). As a result, surgeons become more productive, since they are able to deploy resources more easily to other non-technical aspects of the surgery (Masters et al. 2008, Zhu et al. 2011).

Second, in the acquisition of a cognitive skill, potential errors in initial understanding are gradually detected and eliminated (Fitts 1964). In fact, a widely recognized cognitive theory of

learning (ACT-R adaptive control of thought-rational, Anderson 2013) suggests that a *discrimination process* plays a critical role in the learning and acquisition of a cognitive skill (Anderson 1982). This discrimination process helps one narrow and specify the applicability of new procedural knowledge by producing multiple variants on the conditions of the same action. One important benefit of performing a different and concurrent task is therefore that it facilitates this discrimination process. In facing multiple tasks rather than an individual one, the most appropriate course of action for each particular task is better identified and learned by remembering and comparing its variable bindings (Anderson 1982). Considering all the above arguments, we predict that:

Hypothesis 2: *Concurrent exposure to task variety has a positive impact on subsequent focal task productivity.*

3.5 The Moderating Role of Short-Term Non-Concurrent Variety

We next focus on the way recent (short-term) non-concurrent variety interacts with long-term non-concurrent variety. We argue that the negative effects of non-concurrent variety on surgeons' decision-making skills will be aggravated by recent non-concurrent exposure to variety.

As discussed in Section 3.3, non-concurrent exposure to variety may decrease surgeons' productivity due to their response to similar stimuli, which may prove detrimental over time. Researchers have shown that priming, which is also observed among medical staff (Loewenstein and Lerner 2002, Bargh and Williams 2007), is more likely to happen when exposure to the related task is more recent (Lerner et al. 2004, Bargh and Williams 2006). Compared with more distant experiences, the carryover effects of recent experiences are more likely to cause automatic responses to subsequent similar experiences (Bargh and Williams 2006). The reason is that priming is more likely to occur in the presence or even vestiges of recent relevant behavior (Bargh et al. 2012), since individuals tend to respond unconsciously with their most recent relevant behaviors. We therefore expect that the detrimental effect due to unconscious response (Bargh and Williams 2006) caused by non-concurrent variety will be more significant when a surgeon has recently performed another task independently (non-concurrently).

In addition, the negative knowledge-transfer effect (Gick and Holyoak 1987) and erroneous generalizations of non-current variety will be further amplified by recent non-concurrent exposure to variety. Individuals tend to retrieve schemas that they have used recently, even when more plausible and reasonable alternatives exist (Reder 1982). So, a recent non-concurrent exposure to variety will result in surgeons to adopt its schema, which may not be appropriate for the current task. Also, recent variety may amplify the difference between cognitive perception and the actual level of knowledge, which is detrimental for productivity (Hollyoak 1985, Zollo and Reuer 2010). As a result, the probability of negative knowledge transfer due to misjudgments about the similarity between the focal task and other tasks will be higher. We therefore expect that:

Hypothesis 3: *Recent non-concurrent exposure to variety amplifies the negative effect of non-concurrent exposure to variety on subsequent focal task productivity.*

3.6 The Moderating Role of Short-Term Concurrent Variety

Finally, we consider the moderating role of recent (short-term) concurrent exposure to variety on the relationship between concurrent exposure to task variety and productivity. We expect that the positive knowledge transfer from all past concurrent exposures to variety will be higher in the presence of a recent concurrent variety.

As discussed in Section 3.4, an important way that concurrent variety improves subsequent focal task productivity is "implicit learning", in which variation in the context helps individuals improve both recall and understanding of the focal task (Maskarinec and Thompson 1976). Because this process is essentially about identifying and recalling associations of the focal task (Maskarinec and Thompson 1976), which implicitly involves time, exposure to recent concurrent variety will further improve implicit learning and promote neural efficiency.

In addition, individuals tend to forget (Shtub et al. 1993, Egelman et al. 2016) which can have a significant impact on their productivity in procedural cognitive tasks (Bailey 1989, Nembhard and Uzumeri 2000) such as cardiac operations. This forgetting effect becomes less rapid as the complexity of the performed task is increased (Lance et al. 1998, Nempbhard 2000) while they forget more rapidly when the time interval between two consecutive tasks is increased (Globerson et al. 1989, Bailey 1989) even in a procedural cognitive task (Nembhard and Uzumeri 2000) such as a cardiac operation. Researchers have also shown that learned skills (both technical and cognitive ones) tend to deteriorate for surgeons after some time (Ramdas et al. 2016) and have highlighted the need for periodic remediation of any necessary skills (Kahol et al. 2010). Moreover, it has been suggested that performing multiple surgeries concurrently can increase surgeons' retention of skills (Kahol et al. 2010). Hence we believe that a recent concurrent exposure will significantly reduce the forgetting effect for a surgeon. With reduced forgetting, long-term learning effect from exposure to variety will increase.

Finally, the probability of misguided generalizations will be further decreased when a surgeon recently performs an operation that includes the focal task and another one concurrently. Recent concurrent exposure will allow the surgeon to better understand the differences among different surgery types, which will decrease—if not eliminate—the likelihood of invalid generalizations and the risk of adopting inefficient strategies when subsequently performing the focal task (CABG). Therefore, we expect that:

Hypothesis 4: *Recent concurrent exposure to variety amplifies the positive effect of concurrent exposure to variety on subsequent focal task productivity.*

4. Data and Variables

The organization that we use for our study is the cardiac unit of a private hospital in Europe that is the property of an American non-profit organization. The hospital admits more than 2,000 patients annually and performs around 850 cardiac operations each year. We test our hypotheses using an archival dataset of all 3,275 CABG operations performed in the hospital during the period from 01/01/2004 to 31/03/2011. After removing two operations with missing data, we are left with 3,723 operations for our study.

Each surgery team consists of one lead surgeon and zero to four assistant surgeons. There are 44 surgeons in our sample: 19 started working after the beginning of our dataset, and 13 do not appear during the last year of our dataset. Apart from the surgeons, each team also typically has one anesthesiologist, one perfusionist, and zero to three scrub nurses. We provide further information on surgeons in the Appendix. Like other studies that examine the role of experience on performance in surgical settings (see, for example, KC and Staats 2012), our study concentrates on surgeons (lead and assistant surgeons, 44 in total) and their exposure to various tasks. There are two reasons for this. First, different surgery types will result in surgeons performing different set of steps and activities during the surgery, providing many opportunities for learning. However, tasks and activities for the rest of the team members (e.g., the anaesthesiologist preparing the patient, nurses providing the equipment, etc.) are more trivial and very similar. Since our primary research agenda in this paper is to identify the effect of task variety on learning and productivity, we focus on surgeons, who do perform somewhat different activities and tasks. Second, interviews with the medical staff at the hospital confirmed our intuition that we should consider only surgeon members of the team when studying the role of task variety on productivity in this setting.

Our dataset contains information about the type of operation performed, the members of the surgical team, and the operation's duration, including exact start and end times. Our sample also includes information regarding the patient's condition prior to operation. Specifically, the hospital labels each patient's case either "severe", "medium" or "mild". Finally, our dataset includes information about in-hospital mortality for patients who have had an operation in the hospital.

To test our hypotheses, we use CABG as the focal surgery type and examine the impact of exposure to other cardiac surgery types (i.e., concurrently vs. non-concurrently) on the duration of subsequent CABG-only surgeries. We choose to use CABG as our focal task because it is the most common cardiac surgery type (Clark and Huckman 2012), and indeed it appears more than any other surgery type in our data set. In addition, since our goal is to examine the different effects of concurrent and non-concurrent task variety on subsequent productivity in the individual focal task, we need a task which could be performed both concurrently with another task and on its own. Finally, CABG surgeries have received significant attention in the recent operations management literature (Pisano et al. 2001, Huckman 2003, Huckman and Pisano 2006, KC and Staats 2012), which could help us consider our results' validity and generalizability. Consequently, CABG is an ideal choice to consider as the focal task for our study.

A surgeon can perform multiple surgery types on the same patient during the same operation, as required by medical conditions. For this analysis, in addition to the focal task (i.e., CABG), we have information regarding other types of cardiac surgeries performed during the same time interval in the hospital. There are 1,324 valve repair/replacements, 86 congenital surgeries, 70 heart failure procedures, 20 tumour removals, 185 routine cardiac surgeries, and 78 other normal surgeries (all other surgeries that do not fit into any of the previous categories). In addition, our dataset includes 951 complex operations in which a CABG surgery and one of the other types of surgeries were performed concurrently. Our dataset also includes 170 very complex operations in which a CABG and two of the other types of surgeries were performed concurrently. Because all our hypotheses address focal task productivity, we use 3,273 CABG-only surgeries as our observations to test the hypotheses. However, when calculating our independent variables, we make use of all 6,171 surgeries (CABG only, other type only, and concurrent ones; it is worth noting that apart from the 951 complex operations in which a CABG is included, there are 14 more with two surgeries other than CABG performed concurrently); details appear in Section 4.4.

Finally, we also conducted a limited number of interviews with several staff members at the hospital. These interviews enabled us to better comprehend how CABG and other surgeries are performed and helped us to understand hospital policies and management practices.

4.1 Variables

Dependent Variable. We use duration of CABG operations as our dependent variable. Extensive research in psychology has shown that decrease of the completion time of a performed task is an indicator of learning and increased productivity (Thurstone 1919, Graham and Gagne 1940). In examining performance implications of experience and learning-related issues, operation completion time is a commonly employed dependent variable. For example, in their investigation of the effect of team familiarity, organizational experience, and role experience on team productivity, Reagans et al. (2005) employed surgery duration as the dependent variable. Similarly, other learning-related studies such as Pisano et al. (2001) and Edmondson et al. (2003) used procedure completion time for cardiac surgery as their dependent variable. Similarly, we contend that lower completion time reflects learning and increased productivity for the surgeons and so use it as our dependent variable. In addition, in our semiformal interviews, staff members in our hospital also confirmed that lower completion times usually reflect better clinical outcomes.

While our dataset also included in-hospital mortalities, unlike in large-scale multi-hospital studies, death events were quite infrequent in our setting of only one unit of a single hospital. Consequently, similar to Pisano et al. (2001), we were not able to detect any significant variables that explain variation in mortality rates other than the clinical condition of the patient (i.e., severe, medium, mild). Therefore, in line with our theoretical development, we have decided to keep the scope of our study on productivity performance and use in-hospital mortality as a robustness check to make sure that shorter completion times do not come at the cost of increased mortality rates.

One may argue that shorter completion times may also represent inattention to the clinical outcome of the operation. However, we believe that this is not the case in our setting. First, prior research has shown that shorter completion times decrease the probability of post-surgical infections for cardiac surgeries (Gaynes et al. 2001, Gibbons et al. 2011) and can actually improve the patient's clinical outcome (Pisano et al. 2001). Second, we have empirically investigated the association between duration and in-hospital mortality to check whether decreased durations are associated with increased mortality rates. Specifically, we have created four groups of operations according to their duration (i.e., those less than the 25th percentile, between the 25th and 50th percentile, between the 50th and 75th percentile, and higher than 75th percentile) and conducted a chi-square test to examine the relationship between in-hospital mortality and being in one of these four groups. Our results indicate that patient deaths are not evenly spread across these four groups (p-value = 0.023) and that the number of observed in-hospital deaths is significantly higher in the group with longest durations than in other groups. This suggests that shorter completion times are not associated with worse outcomes and that longer durations are not desirable from a clinical-outcome point of view. This finding is in line with previous medical studies which found that longer surgical durations are associated with higher probabilities of a surgical site infection (SSI) in several surgery types, including cardiac operations (Gaynes et al. 2001, Gibbons et al. 2011).

4.2 Independent Variables

Non-Concurrent Variety. This variable captures all prior days' noncurrent exposure to variety for the surgeons in a team. For each surgeon in a team, we first count the total number of times since the beginning of our dataset that she has performed a surgery that does not include CABG (that is, other type only), up to the current CABG operation. Please also see Table 1 for an illustration of how we calculate this variable for an individual surgeon.

Concurrent Variety. This variable captures all prior days' concurrent exposure to variety for the surgeons in a team. For each surgeon, we first count the total number of concurrent operations she has performed, that is, an operation which involves a CABG surgery and one of the other types of surgeries up to the current (focal) CABG surgery since the beginning of our dataset. Please also see Table 1 for an illustration of how we calculate this variable for an individual surgeon.

Non-Concurrent Variety x Recent Non-Concurrent Variety. We first calculate Recent Non-Concurrent Variety. To capture Recent Non-Concurrent Variety, we calculate for each surgeon the number of surgeries that did not include a CABG operation (that is, other type only) that she has performed during the week before the current CABG surgery. We then multiply this new variable by the Non-Concurrent Variety variable to create the interaction term.

Given our setting, we use one week to capture a short time period (i.e., recent). Clearly, what constitutes a short time period may differ depending on the operational context. Given operational dynamics in our context and the fact that CABG operations last around three to six hours, we think

that one week is an appropriate choice. (See also Staats and Gino 2012 p. 143 for a similar discussion, which suggests that a week may be an appropriate choice for characterizing a short time period with tasks of five hours long). Consequently, in our study, any exposure to variety that takes place within the week prior to the current operation is considered recent. We also performed robustness checks for our choice of one week by considering slightly longer and shorter time periods, and our results remained the same.

Concurrent Variety x Recent Concurrent Variety. We first calculate Recent Concurrent Variety. To capture Recent Concurrent Variety, we calculate for each surgeon the number of concurrent operations (a CABG and another type concurrently) she has performed during the week before the current CABG surgery. We then multiply this new variable with the variable Concurrent Variety to create the interaction term.

4.3 Control Variables

Focal Experience. For each surgeon, we calculate the number of CABG surgeries (either CABG-only or combined with another surgery type) she has conducted up to the current CABG surgery (excluding the current one) since the beginning of our dataset. This way, we capture each surgeon's total experience with the focal surgery type. Please also see Table 1 for an illustration of how we calculate this variable for an individual surgeon.

Team Size. We control for the size of the surgical team (counting all team members). Larger teams might have more access to experience and resources (Reagans et al. 2005), but smaller work-group size is associated with increased team productivity (Gladstein, 1984), since larger teams sometimes face coordination challenges that decrease their productivity (Hackman 2002).

Time Fixed Effect. To control for potential environmental changes in our setting (such as hospital policy or technological advances) and also organizational experience that may influence surgery durations, we include dummy variables indicating the year, month, and day of the week that the operation took place.

Individual Average Experience. We control for the average experience—measured in number of operations—of the team members other than the lead surgeon. For each team member (excluding the lead surgeon), we calculate the number of times she appears in any operation prior to the current one (not including the current one) since the beginning of our dataset. We then take the sum and divide by the number of team members (excluding the lead surgeon).

Indicators for Severity of the Case. As mentioned, the hospital labels each patient's case mild, medium, or severe. The "medium" category appears most often in our sample, so we include two dummy variables in all models: "Severe" and "mild" are both equal to one if the hospital has labeled the patient as such and zero otherwise. We expect "severe" cases to generally last longer than the other two categories, as indicated by table 1 in the Appendix.

4.4 Calculation of Key Independent and Control Variables

Table 1 shows how we calculate Focal Experience, Concurrent Variety, and Non-Concurrent Variety for each surgeon in our sample. As mentioned, one operation may include more than one surgery type. That is, while the most typical case is to perform only one type of surgery (e.g., only a CABG or only a valve replacement), a good number of operations include more than one surgery type (e.g., a CABG plus a valve replacement), depending on the patient's medical requirements. It is also worth noting that when we calculate Non-Concurrent and Concurrent Variety, we removed the corresponding recent varieties to not confound our variables[3].

At time t_1, the surgeon performs an operation that includes a CABG and a valve repair. Then in our first observation at time t_2 (our observations are the operations that include only a CABG surgery), her score for Focal Experience will be equal to 1, since she has conducted one CABG surgery prior to t_2. Similarly, her score for Concurrent Variety will be equal to 1, because she has conducted a valve repair and CABG concurrently prior to t_2. Finally, her score for Non-Concurrent Variety will be equal to 0, since she has not yet conducted any other type of surgery individually (non-concurrently). Then at time t_3 she performs an operation that includes only a valve repair. In our next observation for our study at time t_4, her score for Focal Experience will become 2, because she has conducted two CABG surgeries (at times t_1 and t_3) prior to t_4. Her score for Concurrent Variety will remain 1, and her score for Non-Concurrent Variety will now become 1 because she has conducted an individual valve repair surgery on a patient prior to t_4 (i.e., at time t_3). Table 3 of the Appendix shows the average, standard deviation, and median for all lead surgeons for these variables.

Table 1. Surgeries performed by an individual surgeon over time ($t=t_1, t_2, t_3, t_4$)

Time	Patient	Performed Surgery		Independent Variables			Dependent Variable
		CABG	Valve Repair	Focal Experience	Non-Concurrent Variety	Concurrent Variety	
t_1	Patient 1	√	√	n/a	n/a	n/a	
t_2	Patient 2	√		1	0	1	Observation 1
t_3	Patient 3		√	n/a	n/a	n/a	
t_4	Patient 4	√		2	1	1	Observation 2

5. Empirical Approach and Results

We test our hypotheses using a fixed-effects panel regression with AR(1) disturbance based on Baltagi and Wu (1999). That is, we examine the effect of concurrent and non-concurrent variety for lead surgeon *i* on the duration of CABG operation *j*. Our panel data structure, in which surgeons perform operations over time, poses several challenges. First, there may be unobserved heterogeneity

[3] We also repeat our analysis without removing corresponding recent variety and get very similar results.

between surgeons in areas such as ability, education, or experience and this heterogeneity may bias our results. Second, there is potential serial correlation between operations that were performed by the same lead surgeon close in time. In fact, when we used the Durbin-Watson test, we found that autocorrelation in our sample is likely to be first-order (i.e., AR(1)). Third, we have an unbalanced panel structure with unequally spaced observations over time. Our fixed-effects regression with AR(1) disturbance based on Baltagi and Wu (1999) explicitly takes into account all of these issues (i.e., xtregar command in Stata).

Also, our analyses of the distribution of dependent and continuous independent variables (*ladder* and *gladder* commands in Stata) revealed skewed distributions. Consequently, and in line with our theoretical arguments and previous studies on conventional learning curve models examining completion times (Argote 1999, Reagans et al 2005), we decided to take the logarithm of all these variables. We also confirmed the normality of residuals and checked for heteroscedasticity by using the Breusch-Pagan test (1979), which did not reject the null hypothesis, thereby confirming that heteroscedasticity does not pose a threat to our analyses.

Our model is the following:

$$
\begin{aligned}
ln(duration_{ij}) = {} & \beta_0 + \beta_1 ln(\text{Non-Concurrent Variety}_{ij}) \\
& \beta_2 ln(\text{Concurrent Variety}_{ij}) + \\
& \beta_3 ln(\text{Recent Non-Concurrent Variety}_{ij}) \times ln(\text{Non-Concurrent Variety}_{ij}) + \\
& \beta_4 ln(\text{Recent Concurrent Variety}_{ij}) \times ln(\text{Concurrent Variety}_{ij}) + \\
& \beta_5 ln(\text{Recent Non-Concurrent Variety}_{ij}) + \\
& \beta_6 ln(\text{Recent Concurrent Variety}_{ij}) + \\
& \beta_7 ln(\text{Focal Experience}_{ij}) + \\
& \beta_8 ln(\text{Team Size}_{ij}) + \\
& \beta_9 ln(\text{Individual Average Experience}_{ij}) + \\
& \beta_{10} (\text{Severe}_{ij}) + \\
& \beta_{11} (\text{Mild}_{ij}) + \\
& \alpha_i + \\
& t_j + \\
& u_{ij}\,, \\
& \text{where } u_{ij} = \rho u_{ij-1} + e_{ij}
\end{aligned}
$$

In the above model, α_i represents lead surgeon fixed effects that are included which allows us to control for unobserved time invariant lead surgeon heterogeneity such as education, background etc. and t_j represents time effect indicating the year, month, and day of the week that the operation takes place. In addition, u_{ij} is the serially correlated error term, with $|\rho| < 1$ being the first-order autocorrelation coefficient, and e_{ij} is independent and identically distributed with zero mean and constant variance.

Table 2 shows descriptive statistics and correlations among the variables. Table 4 shows the results for all our hypotheses. Due to the AR(1) covariance structure we employ, the number of observations in the table is equal to 3,261. In Model 1 we include only our control variables. As expected, *focal experience* and *average individual experience* have negative and significant coefficients, suggesting that they reduce completion times, whereas team size increases duration. In addition, compared with the baseline group of medium, *severe* operations take longer and *mild* operations are shorter in duration.

In Model 2, we add our first variable of interest: *Non-Concurrent Variety*. The adjusted R^2 is increased by 2.8%, and an F-test showed that Model 2 is superior to Model 1 ($p < 0.05$). *Non-Concurrent Variety* has a positive and significant coefficient ($p < 0.01$), providing support for our first hypothesis. In Model 3 we add *Concurrent Variety.* The adjusted R^2 is further increased by 8.84%, and an F-test showed that Model 3 is superior to Model 2 ($p < 0.01$). *Concurrent Variety* has a negative and significant coefficient ($p < 0.01$), supporting Hypothesis 2.

In Model 4 we add the interaction terms *Non-Concurrent Variety x Recent Non-Concurrent Variety* and the variable *Recent Non-Concurrent Variety*. The adjusted R^2 is further increased by 1.25%, and an F-test indicated that Model 4 is superior to Model 3 ($p < 0.05$). We also see that *Non-Concurrent Variety x Recent Non-Concurrent Variety* is significant ($p < 0.01$) and positive, providing support for our third Hypothesis. Finally, in Model 5 we include the interaction term *Concurrent Variety x Recent Concurrent Variety* and the variable *Recent Concurrent Variety*. The adjusted R^2 is further increased by 0.62%, and an F-test indicated that Model 5 is superior to Model 4 ($p < 0.05$). *Concurrent Variety x Recent Concurrent Variety* is significant ($p < 0.05$) and negative, providing support for Hypothesis 4.

According to Model 3, an increase of 20% in *Non-Concurrent Variety* increases duration by 5.95% (17.72 minutes), whereas such an increase of *Concurrent Variety* decreases duration by 5.64% (16.80 minutes).

Next, we repeat our analysis using all the surgeons of the team (not just the lead surgeon). Specifically, we include all surgeons of the team when calculating our independent variables and repeat our analysis while controlling for the experience of the other members (in this case, we calculate *Non-Concurrent Variety*, *Concurrent Variety,* and *Focal Experience* using all the surgeons in the team and then taking the average and *Individual Average Experience* using all team members excluding surgeons). Table 3 shows descriptive statistics and correlations among the variables. Table 5 shows the results. In Model 1 we include only our control variables. In Model 2, we add our first variable of interest, *Non-Concurrent Variety*. The adjusted R^2 is increased by 13.99%, and an F-test showed that Model 2 is superior to Model 1 ($p<0.01$). *Non-Concurrent Variety* has a positive and significant coefficient ($p<0.01$), providing support for our first hypothesis. In Model 3 we add *Concurrent Variety.* The adjusted R^2 is further increased by 9.20%, and an F-test showed that Model 3

is superior to Model 2 ($p<0.01$). *Concurrent Variety* has a negative and significant coefficient ($p<0.01$), supporting Hypothesis 2.

In Model 4 we add the interaction terms *Non-Concurrent Variety x Recent Non-Concurrent Variety* and the variable *Recent Non-Concurrent Variety*. The adjusted R^2 is further increased by 1.12%, and an F-test indicated that Model 4 is superior to Model 3 ($p<0.05$). We also see that *Non-Concurrent Variety x Recent Non-Concurrent Variety* is significant and positive ($p<0.01$), providing support for our third hypothesis. Finally, in Model 5 we include the interaction term *Concurrent Variety x Recent Concurrent Variety* and the variable *Recent Concurrent Variety*. The adjusted R^2 is further increased by 1.11%, and an F-test indicated that Model 5 is superior to Model 4 ($p<0.05$). *Concurrent Variety x Recent Concurrent Variety* is significant and negative ($p<0.01$), providing support for Hypothesis 4.

One limitation of our fixed-effects AR(1) regression based on Baltagi and Wu (1999) is that it was not possible to report robust standard errors. The Breusch-Pagan test (1979) revealed no heteroscedasticity, but, nevertheless, we additionally run fixed-effects regression with robust standard errors (*xtreg fe* with *robust* option) and test all our hypotheses. Tables A5 and A6 in the Appendix show the results for our hypotheses using this alternative model. The results for all our hypotheses remain the same in terms of significance and support and are close in terms of coefficients.

We next investigate the economic significance of our main variables of interest. According to Model 3, an increase of 20% in *Non-Concurrent Variety* increases duration by 5.07% (15.10 minutes), whereas such an increase of *Concurrent Variety* decreases duration by 3.69% (10.99 minutes). These findings suggest that not only are the effects of concurrent and non-concurrent variety statistically significant, but their practical effects on surgery completion times are also considerable.

Next, we conduct post hoc plots for Hypotheses 3 and 4 using the lead surgeon and then all surgeons of our teams to examine the moderating effects described by Aiken and West (1991) and Dawson and Richter (2006). We divide our sample into a subset with values above the median for Recent Non-Concurrent Variety and a subset with values below the median for Recent Non-Concurrent Variety and plot Non-Concurrent Variety and Duration (Figures A1 and A2 in Appendix). Notice that we employ logged variables in the figures. We also divide the sample into a subset with values above the median for Recent Concurrent Variety and one with values below Recent Concurrent Variety and plot Concurrent Variety and Duration (Figures A3 and A4 in Appendix). In dividing our sample, we decided to use the median instead of the mean plus one standard deviation and the mean minus one standard deviation because in the latter case our sample was dramatically decreased. Although all plots reveal the moderating effects as proposed in Hypotheses 3 and 4, the effect of the moderation is quite small in terms of economic significance (also note the coefficients of the interaction terms in our models). That is, combined with the main effects of concurrent and non-concurrent task variety, and their short term effects, we observe quite modest moderation effects in terms of practical magnitude in our sample. This is not surprising in our surgery setting, since the

primary drivers of any surgery completion time are, first and foremost, clinical factors. Consequently, we believe that our theoretical insights on moderation effects are useful despite their limited practical applicability in our setting with their small effect sizes. In addition, in other professional service settings where external factors (e.g., the patient's clinical condition) are not as dominant in driving productivity outcomes (such as legal services, consulting, etc.), we expect such moderation effects to be sizeable and of economic significance.

6. Robustness Checks

We perform several additional analyses to examine the robustness of our results and to rule out potential alternative explanations.

6.1 Data and Variables' Operationalization

As in any empirical study, our dataset is limited. A potential concern is that we have no data for surgeries or surgeons prior to the beginning of our dataset, which may influence our results. To address this, we repeat our analysis after excluding different time intervals from the beginning of our dataset. That is, we remove the first 12 months and 24 months, and we calculate our dependent variable and all independent variables using the remaining data (e.g., when we remove 24 months, we calculate surgery durations using only those surgeries that took place between months 24 and 87, and when calculating concurrent exposure to variety, for instance, we similarly use all operations between months 24 to 87). We then repeat our analysis. The results for all our hypotheses remain the same qualitatively. Finally, Lapre and Tsikriktsis (2006) suggest that studies examining a learning effect after a long time since the beginning of the learning curve can eliminate potential bias by using a log-linear model. Consequently, we repeat our analysis for H1 and H2 by using a log-linear model and find full support for all our hypotheses. We therefore believe that missing experience (i.e., missing data prior to the beginning of our dataset) does not pose a threat to our main findings.

We also test the sensitivity of our results by repeating our analysis after removing the first 12 months of our dataset only for the dependent variable (about 14% of our initial observations). That is, while we include our entire dataset for calculating the independent variables, in this case the set of observations (and hence the dependent variable) start after month 12. This way, we investigate the robustness of our results when there is missing data for our main independent variables at the beginning. While the magnitude of effects change as expected, our primary findings remain the same.

We also replaced Average Individual Experience of the rest of the team members with a variable called Average Individual Direct Experience, which captures the number of times each non-surgeon team member has conducted the focal type of operation (CABG) since the beginning of our dataset. Our results remain the same.

One may also be concerned that the effects we observe may only work on single focal operations (CABG-only) and may not hold on non-focused operations that involve more than just CABG. In

order to examine this, we repeat our analysis after changing the way we define our variables. Specifically, we define as focal experience all CABG and valve repair/replacement surgeries (the second most frequent operation type in our sample) and calculate *Focal Experience*, *Non-Concurrent Variety*, and *Concurrent Variety* accordingly. Specifically, we use all operations that include a CABG or a valve repair/replacement when calculating *Focal Experience*, all operations that include neither CABG nor valve repair/replacement when calculating *Non-Concurrent Variety*, and all operations that include both CABG and valve repair/replacement and another type when calculating *Concurrent Variety*. We repeat our analysis and find support for all our hypotheses. This confirms that the effects of non-concurrent and concurrent exposure to variety hold in operations that include more than a CABG.

Finally, we repeat our analysis using all operations (the whole sample) as observations and not just the CABG operations. This is because, if our results pertaining Non-Concurrent Variety change significantly and we observe an exactly opposite effect, then this could imply very different managerial implications for conducting Non-Concurrent Variety. In this new analysis on the whole sample, our results remained similar to those we obtained with the CABG-only sample, though with smaller effect sizes for both Non-concurrent variety and Concurrent variety. Hence, we believe that our results do not depend on the specific choice of focal operation.

6.2 Surgery Assignments

An additional concern could be the possibility that more severe cases might be assigned to more experienced lead surgeons or, similarly, easier cases to less experienced lead surgeons. To deal with this, we first investigate the distribution of severe cases among surgeons and do not observe any patterns. Second, we conduct a chi-square test to ensure that the mild, medium, and severe cases are evenly spread across lead surgeons. The results (Table A4 in Appendix) indicate that there is no difference across lead surgeons in terms of severity of assignments. Next, we repeat our analysis after dropping the most critical cases from our sample, which includes patients that died in the hospital after the operation, and find the same results qualitatively. We also investigated the spread of deaths among surgeons and examined any potential correlations of these deaths with our key variables. We find that these in-hospital deaths are spread quite evenly across surgeons and show no correlation with our key variables other than the "severe" patient condition.

A severe case can be more critical than another severe case, and, similarly, one mild case might be easier than another. Because the hospital does not make these kinds of distinctions within severe cases and within mild cases, we further examined this issue: We repeated our analysis after excluding the severe cases (using only the mild and medium ones). Then we repeated our analysis after excluding only the mild cases from our initial sample (using only the medium and severe ones). Finally, we also repeated our analysis using only the medium cases. In all three cases, while the magnitude of effects

changed considerably as expected, our results remained the same (please refer to the Appendix for corresponding results).

Finally, we further check the relationship between in-hospital mortality and duration. Specifically, we create two groups of observations: The first group includes all the operations that resulted in the death of the patient after the completion of the operation (i.e., post-operation in-hospital mortality, this is the only post-operation mortality information we have available to us). Also please note that all these operations were characterized as severe. The second group includes all the severe operations that did not result in the death of the patient. Please note that in the former group, we do not include the operations that resulted in the death of the patient during the operation since surgery duration would be biased in these cases. We conduct a t-test to compare the average duration of these two groups. Our results indicate that there is no significant difference in the duration of these two groups (p-value = 0.16). Next, we create a more granular matched sample using the Coarsened Exact Matching algorithm described by Iacus et al. (2011) using the Stata command cem. This enables us to compare the duration of the operations in a more closely matched sample. Specifically, we focus only on the severe cases, exclude the cases in which the patient died during the operation and use a dummy variable equal to 1 if the operation resulted in the death of patient after the operation and 0 otherwise as our treatment variable. We also use the covariates of age, gender, lead surgeon and surgery type for our matching process. Our results indicate that 79% of our observations are matched and the L1 statistic-a widely accepted measure of global imbalance comparison (Iacus et al. 2011) is reduced by 19.6% (from 0.449 to 0.361) indicating therefore a significant decrease of imbalance after our matching process. The t-test after the matching process indicates no significant difference in terms of the duration between the operations with in-hospital mortality and with no mortality (p-value > 0.10). This further supports that reduced surgery durations are not associated with higher mortality rates.

6.3 Instrumental Variables Approach

We conducted fixed-effects regressions, which control for all observed and unobserved time-invariant heterogeneity across lead surgeons through a de-meaning process, as well as further analyses of surgery assignments outlined above. However, these analyses might not have fully addressed potential time-varying individual effects which are unobservable to us but may affect lead surgeons' task variety and productivity at the same time.

To address this, we use an instrumental variables approach with 2SLS specification. Specifically, we use the number of public holidays of the country in which the hospital is located and the number of the lead surgeon's vacation days as instruments for surgeons' concurrent and non-concurrent variety. This approach is similar to Lambrecht et al. (2011) which uses vacations and public holidays as instruments for interruption of individuals' adoption process of online banking service. While in our study we are not examining the effects of interruption per se, surgeons' vacations and public holidays do influence their concurrent and non-concurrent experience gaining process. That is, a lead

surgeon's exposure to variety is influenced during holidays and vacations (i.e., the higher the number of days spent in public holidays and vacations, the lower the opportunities for concurrent or non-concurrent learning), however, the number of days spent in these intervals is unrelated to any CABG assignment and selection issues in the hospital. Also, the cumulative number of such days off should not have any influence on surgery durations.

We create two variables for our instruments: *Public Holidays* and *Days of Vacation*. For every operation, we calculate the number of public holidays prior to that operation and take the sum. Finally, since exposure to variety increases as time goes by, when we define our variable *Public Holidays*, we use the number of days that each lead surgeon has been working at the hospital minus the sum of the public holidays up to this point. We expect to observe a positive association between this variable and lead surgeons' exposure to variety.

Regarding *Days of Vacations*, we define vacations if the lead surgeon does not appear in our sample for two weeks or more (we also considered shorter and longer durations and obtained similar results) and then reappears (which means that she is still working at the hospital). Our surgeons work solely for our client hospital, so their absence in our sample does not indicate that they may be working in another hospital or in private practice. So, for every operation, we calculate total days of vacation the lead surgeon has taken up to the current operation. Finally, since exposure to variety increases as time goes, when we define *Days of Vacations*, we use the number of days that each lead surgeon has been working at the hospital minus the number of days she has taken as vacation up to this point. We expect to observe a positive effect of this variable on lead surgeons' exposure to variety.

We next evaluate the suitability and validity of our two instruments *Public Holidays* and *Days of Vacation*. For an instrument to be valid (and therefore the instrumental variable estimators to be asymptotically consistent), it needs to satisfy the relevance and exclusion restriction conditions (Wooldridge 2010). To satisfy relevance, the instrument should be correlated with the endogenous variable. One can empirically verify this by analyzing the first stage regression. For exclusion restriction to be valid, the instrument should not be correlated with the error term in the main equation, conditional on the other observable covariates. It is, however, not possible to test that an instrumental variable meets the exclusion restriction (Angrist and Pischke 2009, Wooldridge 2010). So, this condition has to be argued qualitatively based on institutional (contextual) knowledge. Consequently, we first empirically verify relevance of our instruments and then explain why we believe exclusion restriction is not violated in our setting.

First, we verified the relevance condition. The coefficient of both Public Holidays and Days of Vacation are always significant and positive in the first stage regressions on Concurrent variety and Nonconcurrent variety with a high F-statistic and high adjusted R^2 of the corresponding models. Specifically, Tables 6, 7, and 8 show the results of our IV approach. In Table 6, we only consider *Non-Concurrent Variety* and use *Public Holidays* as the instrument. Our first-stage estimation in

Model 1 reveals a significant positive effect ($p < 0.01$) of *Public Holidays* on *Non-Concurrent Variety* with an F-statistic well above the common threshold of 10. In Table 7, we focus only on *Concurrent Variety* and use *Public Holidays* as the instrument. We examine the first stage estimation with *Public Holidays* as the instrument and observe a significant ($p < 0.01$) positive effect of *Public Holidays* on *Concurrent Variety*. The F-statistic is also well above the common threshold of 10. In Table 8, we use both *Public Holidays* and *Days of Vacations* as instruments to *Non-Concurrent Variety* and *Concurrent Variety* in two stage estimation. As before, our instruments are significant in the first stage regression with an F-statistic well above 10.

In addition, in all of the above models, the value of the Kleibergen-Paap Wald statistic for weak instruments and the underidentification test (Kleibergen LM statistic) indicated that our instruments indeed have power and satisfy the relevance condition[4].

Second, we explain why *Public Holidays* and *Days of Vacation* do not violate exclusion restriction (i.e., why they are not correlated with the error). For any given operation j by surgeon i, the number of days the surgeon had spent in public holidays or in vacation days from the beginning of our dataset until the focal operation j is unlikely to have any direct effect on the duration of the focal operation j he performs. Notice that, like in many other hospitals, our surgeons' are on call during public holidays on a rotation basis, so surgeons' likelihood of performing an urgent operation during a public holiday is very similar. That is, our instruments *Public Holidays* and *Days of Vacation* are unrelated to any potential unobserved surgeon selection issues, and therefore, it is very unlikely for our instruments to be correlated with the error term.

In addition, the only way our instruments can affect our dependent variable is through the fact that they may be a proxy for 'experience'. But in our case, the experience of a surgeon is comprised of focal, concurrent and non-concurrent experience. Notice that we already control for focal experience in our main (second stage) model, and concurrent and non-concurrent experience are our potentially endogenous variables. Hence, the fact that our instruments can be a proxy for experience is not a problem, because after controlling for focal experience at the second stage, our instruments *Public Holidays* and *Days of Vacation* would not be correlated with the error term in this second stage equation. Thus, it is unlikely that our instruments *Public Holidays* and *Days of Vacation* have any direct relationship with surgery duration after controlling for focal experience. To illustrate, a surgeon's days spent in public holidays or vacations, for example, from January 2004 to January 2006 should have no direct effect on the duration of a surgery he performs in February 2006 after controlling for his Focal experience. Therefore, exclusion restriction is satisfied. These same days, however, will have an effect on his concurrent and nonconcurrent variety scores because the higher the time a surgeon spends on public holidays or in vacations days, the lower the number of opportunities for concurrent and nonconcurrent variety. This satisfies the relevance condition.

[4] Please note that because we have robust standard errors, we used these two tests instead of the Crag-Donald Wald F statistic and the Anderson canonical correlation LM statistic.

Also, notice that the only case where some form of empirical validation for exclusion restriction is possible (though with some strong assumptions) is when the model is overidentified (i.e., there are more instruments than endogenous variables). In our case, we also repeat our analysis using two instruments (both Public Holidays and Days of Vacation) for one endogenous variable (for only Concurrent variety or for only Nonconcurrent variety) in Tables 6 and 7. Hence, we are able to conduct the Sargan overidentifying restriction tests. Our results on both cases reveal that we fail to reject the null hypotheses ($p > 0.10$) that the error term is uncorrelated with the instruments. This supports our previous conclusions.

After discussing the validity of our instruments, we next focus on our second stage estimation results with the instrumental variables approach. Second-stage estimation results in Models 2 and 4 in Tables 6 provide support for our first hypothesis: *Non-Concurrent Variety* has a positive and significant coefficient ($p < 0.01$). Similarly, second-stage results in Model 2 and Model 4 of Table 7 provide support for our second hypothesis: *Concurrent Variety* has a negative and significant coefficient ($p < 0.01$). In addition, second stage estimation (Model 3) in Table 8 provides support for both H1 and H2. Specifically, *Non-Concurrent Variety* has a positive and significant coefficient ($p < 0.01$) and *Concurrent Variety* has a negative and significant coefficient ($p < 0.01$). These results are similar to our previous findings without the use of instruments.

Notice that we used our instrumental variables approach and report 2SLS results as a robustness check to our main model. This is because we do not reject the null hypothesis that our key variables of interest Concurrent variety and Nonconcurrent variety are exogenous. Specifically, we conduct the Durbin-Wu-Hausman test (Durbin 1954, Wu 1973, Hausman 1978) which is defined as the difference between two Sargan-Hansen statistics (one from the OLS regression and one from the IV regression). In other words, the test compares the coefficients between the instrumented regression and the regular OLS regression and under the null hypothesis the endogenous variables can be treated as exogenous. Our result ($p > 0.10$) fails to reject the null hypothesis, suggesting that endogeneity is not a major threat to our original findings. That is, our 2SLS results are not significantly different from the original findings in the main model, we therefore report those more efficient ones in the main model and include 2SLS analyses as a robustness check.

7. Discussion and Conclusions

Professional service firms are the epitome of the increasingly knowledge-based economies of the world. While there is growing interest in the study of professional services in the management literature (Maister 1993, Hinings and Leblebici 2003, Greenwood et al. 2005, Gardner et al. 2008), the study of professional service work from an operations standpoint has been quite limited (Roth and Menor 2003, Lewis and Brown 2012). Despite the clear importance of such white-collar professions to the economies of developed countries, their operations are much less understood than the operations of blue-collar work (Hopp et al. 2009).

Learning is a critical component of white-collar work (Argote and Ingram 2000). From an operations perspective, the essential issue with regards to learning in white-collar settings is how it affects performance (Hopp et al. 2007). While past experience is generally associated with an increased learning rate at the individual level (Narayanan et al. 2009, KC and Staats 2012, Staats and Gino 2012), some experiences can also have a negative effect on individual productivity (Allport et al. 1994, Lapré and Nembhard 2010, Argote and Miron-Spektor 2011, Lapré 2011). Our goal in this paper was to explore the effects of a specific kind of experience—that is, experience from performing related tasks (i.e., related variation)—on individual learning and focal productivity when the related tasks are performed concurrently vs. non-concurrently with the focal task. We find that while concurrently performing another task over time enhances the productivity of the focal task, performing another task non-concurrently reduces this productivity. We also introduce recent exposure to variety as an important moderator which amplifies the relationship between variety and productivity.

While the extant literature has recognised task variety as a potentially important driver of individuals' task performance, evidence on productivity implications of task variety have been mixed. It is only recently that researchers have started to disentangle more nuanced elements of task variety and investigate their effects on individual productivity. In an experimental study, Schilling et al. (2003) make an important distinction between related and unrelated task variation and show that the learning rate under conditions of related variation is significantly greater than under conditions of unrelated variation. In two other recent studies, Staats and Gino (2012) investigate the productivity implications of exposure to variety in the long term vs. in the short term, and KC and Staats (2012) introduce subtask variety (within task variety) as an important driver of individual performance. Although these studies have considerably improved our understanding of task variety's effects on individual productivity, several elements of task variety, such as the relationship between focal tasks and varied tasks and their impact on productivity, have been much less well understood. Our study sought to contribute to this line of inquiry by focusing on when the varied task has been performed with regards to the focal task. That is, we introduce the way different task are performed (concurrently vs. non-concurrently) as an important dimension to consider in understanding the effect of task variety on productivity.

Our extensive dataset, which covers a time interval of more than seven years, allowed us to investigate the influence of exposure to variety on individual learning and productivity over time. In addition, because we are able to observe a variety of surgeries performed in different configurations (e.g., CABG only, other type only, CABG and other type) which are driven by exogenous (i.e., medical) factors, our results are not likely to be affected by endogeneity concerns.

Surgeons have been identified as typical 21st-century knowledge workers (KC and Staats 2012) in that they experience constant learning throughout their careers. Accordingly, our results can provide useful insights for other professional service workers with regards to task allocation and timing

strategies. Our findings are particularly relevant to settings that are characterized by high levels of worker discretion and control over how to conduct a variety of tasks. Our results suggest that performing other related tasks concurrently with a focal common task can improve individuals' learning and productivity over time. Our results indicate that a 20% increase in exposure to other related tasks can decrease the time an individual needs to perform the focal task by 3.69% (which, in the case of our surgeons, translates into around 17 CABG operations per year in terms of average time savings). On the other hand, we observe that performing related tasks in isolation (non-concurrently) does not provide such learning and productivity improvement opportunities. In contrast, such non-concurrent variety could be detrimental to productivity due to negative transfer effects. Our results suggest that a 20% increase in non-concurrent exposure to variety can decrease average surgery duration by 5.07% (which, for our surgeons, translates to around 23 CABG operations per year in terms of average time savings). To maximize individual learning and improve productivity in common tasks, therefore, knowledge workers may consider pairing their most common task(s) with other related tasks and try performing them concurrently as much as possible. In addition, individuals should pay close attention to related tasks which may appear similar at surface but may have nuanced structural differences as negative knowledge transfer and erroneous generalizations may occur between such tasks, especially when they are performed at different times (non-concurrently). In addition, our results on the moderating role of recent exposure to variety suggest that short-term exposure to variety may matter for subsequent task productivity. That is, recent variety amplifies the respective influences of concurrent and non-concurrent variety on subsequent task productivity. Therefore, when individuals are devising strategies to improve their productivity on tasks they perform frequently, it may help to consider tasks that they have carried out both in the long term and recently.

As in all empirical studies, our study has its limitations. First, although we have highly granular longitudinal data on surgery operations in terms of their type, constituents, duration, surgeon team, among others, we have limited information about patients' condition before the operation. Specifically, our hospital groups patients into mild, medium, and severe categories by taking into account various clinical factors such as patient's age, other medical conditions, and surgery requirement, and this is the only patient related information available to us. Admittedly, this is quite a crude characterization, and there is likely variation within the three groups (e.g., some cases might be more severe than others despite being in the same severe category). Ideally, we would like to have more detailed information about patients' condition, such as their EuroScore or Higgins score (KC et al. 2013), but this was not available in our dataset. In examining the role of task variety, future work could combine detailed surgery data with more granular patient characteristics to develop more comprehensive risk and outcome measures for patients.

A related issue is that, although we use above categories as indicator variables, control for the surgery type and also run multiple robustness checks in various more homogenous subsamples (such

as only medium cases etc.), we cannot control for all unobserved patient heterogeneity. If such unobserved patient heterogeneity and our variables of interest such as concurrent and non-concurrent variety are correlated, this would then bias our estimates. We addressed these concerns by examining hospital's patient allocation process qualitatively through interviews and empirically (which do not indicate any systematic bias of assigning certain cases to certain surgeons), by including surgeon fixed effects, and by conducting an instrumental variables analysis. Also, in our study, observations (i.e., CABG operations) were among the most homogenous surgery types, alleviating to a degree, concerns on unobserved heterogeneity. Future research could examine performance implications of concurrent variety and non-concurrent variety in settings where individuals are assigned to cases or they are exposed to variety in a naturally random manner (e.g. through rotations, fully randomized assignment of tasks by design, etc.). Future research could also consider exploiting policy changes (via quasi-natural experiment designs) which may result in certain tasks that have been previously done concurrently to move to be conducted non-concurrently, or vice versa. This would provide a cleaner setting to examine potential contrasting effects of concurrent and non-concurrent variety on performance.

In addition, considering the importance of completion time as a key operational performance measure and availability of our detailed surgical duration data, we focus on productivity implications of concurrent and non-concurrent variety. We used our limited in-hospital mortality data to ensure that shorter duration times do not come at the expense of higher mortality rates[5]. However, examining quality implications of concurrent and non-concurrent exposure to variety or other task configurations are clearly appealing future research directions. After all, if certain task configurations provide better learning opportunities and result in higher quality outcomes for knowledge workers, identifying and examining these would provide important contributions to operations management literature.

Although surgeons are characterized as a good example of 21st century knowledge workers in the extant literature and cardiac operations provided us a very suitable setting to study the role of task variety in highly cognitive and knowledge intensive yet repetitive tasks, surgery settings could be quite unique in terms of nature of tasks, life-death outcomes, and learning mechanisms. So, one has to be cautious in generalizing our findings to other professional service settings. It would be interesting to see future research on this domain, and whether and to what extent our contrasting results on concurrent and non-concurrent variety would hold in other professional service settings such as law, consulting, R&D, among others.

Finally, our focus in this study has been productivity implications of surgeons' learning from concurrent and non-concurrent task variety. As such, we are not interested in the entire surgical team but rather individual surgeons and how they learn from task variety concurrently versus non-concurrently. This is because while different surgery types will result in surgeons to perform different

[5] Unfortunately, we also do not have other quality measures such as readmission rates available to us.

set of steps and procedures during the surgery; tasks and activities for other team members (e.g., anaesthesiologist preparing the patient, nurses providing the equipment, etc.) are more or less the same in all surgery types. So getting exposure to task variety only makes sense for surgeons. However, learning is a complex phenomenon with many individual, team-level, contextual and organizational dynamics. Considering the recent research emphasis on the role of team-level mechanisms such as the role of team experience on learning and performance (Reagans et al. 2005, Huckman and Staats 2011, Avgerinos and Gokpinar 2016), future research could explore the interplay between task and team configurations, and their learning implications in less hierarchical team settings with less clear task boundaries such as management consulting or product development teams.

References

Aiken, L. S., S. G. West. 1991. *Multiple regression: Testing and interpreting interactions*. Sage, Thousand Oaks, CA.

Allport D. A., G. Wylie. 1999. Task-switching: Positive and negative priming of task-set. G.W. Humphreys, J. Duncan, A.M. Treisman eds. *Attention, Space and Action: Studies in Cognitive Neuroscience*. Oxford University Press, Oxford, 273-296.

Allport, D. A., E. A. Styles, S. Hsieh. 1994. Shifting intentional set: Exploring the dynamic control of tasks. C. Umilta, M. Moscovitch, eds. *Attention and Performance XV: Conscious and Non-conscious Information Processing*. MIT Press, Cambridge, MA, 421–452.

Anderson, J. R. 2013. *The architecture of cognition*. Psychology Press, New York, NY.

Anderson, J. R. 1982. Acquisition of cognitive skill. *Psych. Rev*. 89(4) 369.

Angrist J.D., Pischke J.S. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press, Princeton, NJ.

Argote, L. 1999. *Organizational learning: Creating, Retaining and Transferring Knowledge*. Kluwer Academic Publishers, Boston, MA.

Argote, L., E. Miron-Spektor. 2011. Organizational Learning: From Experience to Knowledge. *Organ. Sci.* 22(5) 1123-1137.

Argote, L., B. McEvily, R. Reagans. 2003. Managing knowledge in organizations: An integrative framework and review of emerging themes. *Management Sci.*49(4) 571-582.

Argote, L., P. Ingram. 2000. Knowledge transfer: A basis for competitive advantage in firms. *Organ. Behav. Human Decision Processes* 82(1) 150–169.

Avgerinos E., B. Gokpinar. 2016. Team Familiarity and Productivity in Cardiac Surgery Operations: The Effect of Dispersion, Bottlenecks, and Task Complexity. *Manufacturing & Service Oper. Management. Forthcoming.*

Bailey, C.D. 1989. Forgetting and the learning curve: A laboratory study. *Management Sci.* 35(3) 340-352.

Baltagi, B. H., and P. X. Wu. 1999. Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory* 15: 814–823.

Bargh, J. A., E. L. Williams. 2006. The automaticity of social life. *Current Directions in Psych. Sci.* 15(1) 1-4.

Bargh, J. A., E. L. Williams. 2007. On the automatic or nonconscious regulation of emotions. J. Gross eds. *Handbook of Emotion Regulation*. Guilford Press, NY, 429-445.

Bargh, J. A., K. L. Schwader, S. E. Hailey, R. L. Dyer, E. J. Boothby. 2012. Automaticity in social-cognitive processes. *Trends in Cognitive Science.* 16(12) 593-605.

Boh, W. F., S. A. Slaughter, J. A. Espinosa. 2007. Learning from experience in software development: A multilevel analysis. *Management Sci.* 53(8) 1315–1331.

Bongers, P. J., D. P. van Hove, L. P. Stassen, J. Dankelman, H. W. Schreuder. 2015. A new virtual-reality training module for laparoscopic surgical skills and equipment handling: Can multitasking be trained? A randomized controlled trial. *J. of Surgical Education.* 72(2) 184-191.

Bransford J.D., J.J. Franks. 1976. Toward a framework of understanding learning. *Psychology of Learning and Motivation.* 10 93-127

Breusch, T. S., A. R. Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: J. of the Econom. Soc.* 47 1287-1294.

Cellier, J. M., H. Eyrolle. 1992. Interference between switched tasks. *Ergonomics* 35(1) 25–36.

Ch'ng, S. L., A. D. Cochrane, R. Wolfe, C. Reid, C. I. Smith, J. A. Smith. 2015. Procedure-specific cardiac surgen volume associated with patient outcome following valve surgery, but not isolated CABG surgery. *Heart, Lung and Circulation*. 24(6) 583-589.

Clark, J. R., R. Huckman. 2012. Broadening focus: Spillovers, complementarities and specialization in the hospital industry. *Management Sci.* 58(4) 708-722.

Cohen, M. D., P. Bacdayan. 1994. Organizational routines are stored as procedural memory: Evidence from a laboratory study. *Organ. Sci.* 5(4) 554-568.

Craik, F.I. M., E. Tulving. 1975. Depth of processing and the retention of words in episodic memory. *J. of Exper. Psych.* 104(3) 268-294.

Dawson, J. F., A. W. Richter. 2006. Probing three-way interactions in moderated multiple regression: development and application of a slope difference test. *J. of Appl. Psych.* 91(4) 917-926.

Durbin J. 1954. Errors in Variables. *Rev. of the Inter. Stat. Institute.* 22(1/3) 23-32.

Edmondson, A. C. 1999. Psychological safety and learning behavior in work teams. *Admin. Sci. Quart.* 44(2) 350-383.

Edmondson, A. C., A. B. Winslow, R. M. J. Bohmer, G. P. Pisano. 2003. Learning how and learning what: Effects of tacit and codified knowledge on performance improvement following technology adoption. *Decision Sci*. 34(2) 197-224.

Egelman C.D, D. Epple, L. Argote, E.R.H. Fuchs. 2016. Learning by Doing in Multiproduct Manufacturing: Variety, Customizations, and Overlapping Product Generations. *Management Sci.* Articles in Advance. 1-19.

Fitts, P. M. 1964. Perceptual-Motor Skill Learning. A. W. Melton. *Categories of Human Learning*. Academic Press, New York, NY.

Gardner, H. K., Anand, N., & Morris, T. 2008. Chartering new territory: Diversification, legitimacy, and practice area creation in professional service firms. *Journal of Organizational Behavior*. 29 1101–1121.

Gaynes, R. P., D. H. Culver, T. C. Horan, J. R. Edwards, C. Richards, J. S. Tolson, and the National Nosocomial Infections Surveillance System. 2001. Surgical site infections (SSI) rates in the United States, 1992-1998: The national nosocomial infections surveillance system basic SSI risk index. *Clinical Infectious Diseases*. 33(Suppl 2) 69-77.

Gibbons, C., J. Bruce, J. Carpenter, A. P. Wilson, J. Wilson, A. Pearson, D. L. Lamping, Z. H. Krukowski, B. C. Reeves. 2011. Identifications of risk factors by systematic review and development of risk-adjusted models for surgical site infection. *Health Technological Assessment,* 15(30) 1-156.

Gick, M. L., K. L. Holyoak. 1987. The cognitive basis of knowledge transfer. S. M. Cormier, J. D. Hagman eds. *Transfer of Learning: Contemporary Research and Applications.* Academic Press, New York.

Gladstein, D. L. 1984. Groups in context: A model of task group effectiveness. Admin. Sci. Quart. 29(4) 499-517.

Globerson, S., N. Levin, A. Shtub. 1989. The impact of breaks on forgetting when performing a repetitive task. *IIE Transactions.* 21(4) 376-381.

Graham, C., R. Gagne. 1940. The acquisition, extinction, and spontaneous recovery of a conditioned operant response. *J. of Experiment. Psych.* 26(3) 251-280.

Greenwood, R., Li, S. X., Prakash, R., & Deephouse, D. L. 2005. Reputation, diversification, and organizational explanations of performance in professional service firms. *Organization Science*, 16 661–673.

Grierson, L., M. Melnyk, N. Jowlett, D. Backstein, A. Dubrowski. 2011. Bench model surgical skill training improves novice ability to multitask: A randomized controlled study. J. D. Wetswood, S.W. Westwood, L. Fellander-Tsai, R.S. Haluck, H.M. Hoffman, R.A. Robb, S. Senger, K.G. Vosburgh eds. *Medicine meets virtual reality 18.* IOS Press, Netherlands, 192-198.

Hackman, J. R. 2002. *Leading Teams: Setting the Stage for Great Performances*. Harvard Business Press, Boston, MA.

Hackman, J. R., G. R. Oldham. 1976. Motivation through the design of work: Test of a theory. *Organ. Behav. and Human Performance*. 16(2) 250-279.

Haleblian, J., S. Finkelstein. 1999. The influence of organizational acquisition experience on acquisition performance: A behavioral learning perspective. *Admin. Sci. Quart.* 44(1) 29-56.

Hausman J. A. 1978. Specification tests in econometrics. *Econometrica*. 46(6) 1251-1271.

Herzberg, F. 1968. One more time: How do you motivate employees? *Harvard Bus. Rev.* 46(1) 53–62.

Hinings, C. R., & Leblebici, H. 2003. Introduction: Knowledge and professional organizations. *Organization Studies*. 24 827–830.

Holyoak, K. J. 1985. The pragmatics of analogical transfer. G. H. Bower eds. *The Psychology of Learning and Motivation*, Vol. 19. Academic Press, New York, 59-87.

Holyoak K.J, K. Koh. 1987. Surface and structural similarity in analogical transfer. *Memory and Cognition*. 15(4) 332-340.

Hopp, W. J., S. M. R. Iravani, F. Liu. 2009. Managing white-collar work: An operations-oriented survey. *Production and Oper. Management*. 18(1) 1-32.

Hopp, W. J., S. M. R. Iravani, G. Yuen. 2007. Discretionary task completion: A key difference between white-collar and blue collar work systems. *Management. Sci.* 53(1): 61–77.

Huckman, R. S. 2003. The utilization of competing technologies within the firm: Evidence from cardiac procedures. *Management Sci.* 49(5) 599–617.

Huckman, R. S., G. P. Pisano. 2006. The firm specificity of individual performance: Evidence from cardiac surgery. *Management Sci.* 52(4) 473–488.

Iacus S. M., King G., Porro G. 2011. Multivariate matching methods that are monotonic imbalance bounding. *J. of the American Statistical Association* 493(106) 345-361.

Ibanez M., J. R. Clark, R. S. Huckman, B. R. Staats. 2015. Discretionary Task Ordering: Queue Management in Radiological Services. Harvard Business School Technology & Oper. Management Unit Working Paper No. 16-051.

Kahol, K., A. Ashby, M. Smith, J. Ferrara. 2010. Quantitative evaluation of retention of surgical skills in simulation. *J. of Surgical Education.* 67(6) 421-426.

KC, D. S., BR. Staats. 2012. Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. *Manufacturing & Service Oper. Management.* 14(4) 618-633.

KC, D.S., B.R. Staats, F. Gino. 2013. Learning from My Success and from Others' Failure: Evidence from Minimally Invasive Cardiac Surgery. *Management Sci.* 59(11) 2435-2449.

KC, D.S., 2014. Does Multitasking Improve Performance? Evidence from the Emergency Department. *Manufacturing & Service Oper. Management.* 16(2) 168–183.

Lambrecht, A., Seim K.,Tucker C. 2011. Stuck in the adoption funnel: The effect of interruptions in the adoption process on usage. *Marketing Sci.* 30(2) 355-367.

Lance, C. E., A. G. Parisi, W. R. Bennett, M. S. Teachout, D. L. Harville, M. L. Welles. 1998. Moderators of skill retention interval/performance decrement relationships in eight U.S. Air Force enlisted specialties. *Human Performance*. 11(1) 103-123.

Langer, E. J. 1989. *Mindfulness*. Addison-Wesley Publications, Reading, MA.

Lapré, M. A. 2011. Inside the learning curve: Opening the black box of the learning curve. Jaber M. Y., eds. *Learning Curves: Theory, Models, and Applications.* CRC Press, Boca Raton, FL, 23–35.

Lapré, M. A., I. M. Nembhard. 2010. Inside the organizational learning curve: Understanding the organizational learning process. *Foundations and Trends in Tech., Inform. and Oper. Management* 4(1) 1-103.

Lapré M. A., Tsikriktsis N. 2006. Organizational learning curves for customer dissatisfaction: Heterogeneity across airlines. *Management Sci.* 52(3) 352–366.

Lee, T. D., D. Simon. 2004. Contextual Interference. Williams, A. M., N. J. Hodgeseds eds. *Skill Acquisition in Sport: Research, Theory and Practice*. Routledge, London, 29-44.

Lerner, J. S., D. A. Small, G. Loewenstein. 2004. Heart strings and purse strings: Carry-over effects of emotions on economic transactions. *Psych. Science.* 15 337-341.

Levinthal, D. Rerup C. 2006. Crossing an apparent chasm: Bridging mindful and less-mindful perspectives on organizational learning. *Organ. Sci.* 17(4) 502–513.

Lewis, M. A., A. D. Brown. 2012. How different is professional service operations management? *J. of Oper. Management* 30(2) 1-11.

Loewenstein, G., J. S. Lerner. 2002. The role of affect in decision making. Davidson, R., K. Scherer, H. Goldsmith eds. *Handbook of Affective Science*. Oxford University Press, NY. 619-642.

Maister, D. H. 1993. *Managing the Professional Service Firm*, The Free Press, NY.

Maskarinec, A. S., C. P. Thompson. 1976. The within-list distributed practice effect: Tests of the varied context and varied encoding hypothesis. *Memory and Cognition* 4 741–746.

Masters, R. S. W., C. Y. Lo, J. P. Maxwell, N. G. Patil. 2008. Implicit motor learning in surgery: implications for multi-tasking. *Surgery.* 143(1) 140-145.

McKinsey Global institute. 2012. The social economy: Unlocking value and productivity through social technologies. November 2012.

Monsell, S. 2003. Task switching. *Trends in Cognitive Sci*. 7(3) 134-140.

Narayanan, S., S. Balasubramanian, J. M. Swaminathan. 2009. A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management Sci*. 55(11) 1861-1876.

Nembhard, D. A. 2000. The effect of task complexity on learning and forgetting: A field study. *Human Factors.* 42(2) 272-286.

Nembhard, D. A., M. V. Uzumeri. 2000. Experiential learning and forgetting for manual and cognitive tasks. *International J. of Industrial Ergonomics.* 25 315-326.

Nembhard, I. M., A. C. Edmondson. 2006. Making it safe: The effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *J. of Organ. Behav*. 27(7) 941-966.

Nokes, T.J. 2009. Mechanisms of knowledge transfer. *Thinking and Reasoning*. 15(1) 1-36.

Novick, L.R. 1988. Analogical transfer, problem similarity, and experitise. *J. of Exper. Psych.: Learning, Memory, and Cognition.* 14(3) 510-520.

Pisano, G. P., R. M. J. Bohmer, A. C. Edmondson. 2001. Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Sci*. 47(6) 752-768.

PwC. 2012. What are professional services? Retrieved August 15, 2014, http://www.pwc.co.uk.

Ramdas, K., K. Saleh, S. Stern, Liux H. 2016. Variety and experience: Learning and forgetting in the use of surgical devices. *Management Sci.* Forthcoming.

Reagans, R., L. Argote, D. Brooks. 2005. Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Sci*. 51(6) 869-881.

Reber, A. 1989. Implicit learning and tacit knowledge. *J. Experiment. Psych.* 118 219–235.

Reder, L. M. 1982. Plausibility judgements versus fact retrieval: Alternative strategies for sentence verification. *Psych. Rev.* 19 90-137.

Reznick, R. K., H. MacRae. 2006. Teaching surgical skills-changes in the wind. *The New England J. of Medicine*. 355 2664-2669.

Roth, A. V., & Menor, L. J. 2003. Insights into service operations management: a research agenda. *Production and Operations Management.* 12(2), 145-164.

Schilling, M. A., P. Vidal, R. E. Ployhart, A. Marangoni. 2003. Learning by doing something else: Variation, relatedness, and the learning curve. *Management Sci*. 49(1) 39-56.

Shea, J. B., S. T., Zimny. 1983. Context effects in memory and learning movement information. R. H. Magil eds. *Memory and Control of Action*. North Holland, Netherlands, 345-366.

Shtub A., N. Levin, S. Globerson. 1993. Learning and forgetting industrial skills: An experimental model. *Inter. J. of Human Factors in Manufacturing.* 3(3) 293-305.

Spencer, F.C. 1978. Teaching and measuring surgical techniques-the technical evaluation of competence. *Bull. of the Amer. College of Surgeons* 63 9-12.

Staats, BR., F. Gino. 2012. Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Sci*. 58(6) 1141-1159.

Thurstone, L. L. 1919. The learning curve equation. *Psych. Monographs: General and Appl.* 26(3) i-51.

Tucker, A. L., A. C. Edmondson. 2003. Why hospitals don't learn from failures: organizational and psychological dynamics that inhibit system change. *California Management Rev*. 45(2) 55-72.

Tucker, A. L., I. M. Nembhard, A. C. Edmondson. 2007. Implementing new practices: An empirical study of organizational learning in hospital intensive care units. *Management Sci*. 53(6) 894-907.

Tulving, E., D.M. Thomson. 1973. Encoding specificity and retrieval processes in episodic memory. *Psych. Rev.* Vol 80(5) 352-373.

Tversky, A. 1977. Features of similarity. *Psych. Rev*. 84(4) 327.

Von Nordenflycht, A. 2010. What is a professional service firm? Toward a theory and taxonomy of knowledge-intensive firms. *Acad. of Management Rev.* 35(1) 155-174.

Waszak, F., B. Hommel, A. Allport. 2003. Task-switching and long-term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psych.* 46(4) 361-413.

Wooldridge, J.M., 2010. *Econometric analysis of cross section and panel data*. MIT press. Cambridge, MA.

Wetzel, C. M., R. L. Kneebone, M. Woloshynowych, D. Nestel, K. Moorthy, J. Kidd, A. Darzi. 2006. The effects of stress on surgical performance. *The Amer. J. of Surgery*. 191(1) 5-10.

Wu D. M. 1973. Alternative tests of independence between stochastic regressors and disturbances. *Econometrica* 41(4) 733-750.

Wulf, G., R. A. Schmidt. 1997. Variability of practice and implicit motor learning. *J. Experiment. Psych.: Learn. Memory, and Cognition*. 23 987–1006.

Zhu, F. F., J. M. Poolton, M. R. Wilson, Y. Hu, J. P. Maxwell, R. S. W. Masters. 2011. Implicit motor learning promotes neural efficiency during laparoscopy. *Surgical Endoscopy.* 25(9) 2950-2955.

Zollo, M., J. J. Reuer. 2010. Experience spillovers across corporate development activities. *Organ. Sci*. 21(6) 1195-1212.

Table 2. Descriptive Statistics for Lead Surgeons

Variable	Mean	Std. Dev.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12
1. Duration	5.681	0.176	4.094	6.579	1											
2. Recent Non-Concurrent Variety	0.741	0.633	0	1.386	0.011**	1										
3. Non-Concurrent Variety	4.999	0.983	0	5.820	0.121*	0.598**	1									
4. Recent Concurrent Variety	0.505	0.601	0	1.792	-0.019*	0.494**	0.274**	1								
5. Concurrent Variety	4.904	0.984	0	5.489	0.092**	0.487**	0.584**	0.208**	1							
6. Recent Non-Concurrent Variety x Non-Concurrent Variety	4.827	1.954	0	7.077	0.023*	0.828**	0.611**	0.342**	0.611**	1						
7. Recent Concurrent Variety x Concurrent Variety	4.184	1.997	0	6.956	-0.037*	0.389**	0.447**	0.884**	0.471**	0.558**	1					
8. Focal Experience	5.978	1.106	0	6.477	0.052**	0.354**	0.578**	0.254**	0.689**	0.421**	0.289**	1				
9. Team Size	1.617	0.130	0.693	2.079	0.117**	-0.097*	-0.043**	-0.016	0.052*	-0.311**	-0.273	0.049**	1			
10. Average Individual Experience	6.085	1.100	0	7.701	0.103**	0.101*	0.533**	0.041**	0.623**	0.477**	0.293**	0.686**	0.069**	1		
11. Severe	0.043	0.203	0	1	0.036*	0.035*	0.011*	-0.013	-0.061**	-0.013	-0.008	-0.048**	0.024	-0.064**	1	
12. Mild	0.291	0.454	0	1	-0.061**	-0.011	-0.014+	-0.018	-0.029	-0.011	0.005	-0.035*	-0.024	-0.047**	-0.136**	1

+, * and ** denote significance at 10%, 5% and 1% levels respectively

Logged values of all variables except Severe and Mild

Table 3. Descriptive Statistics for all Surgeons

Variable	Mean	Std. Dev.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12
1. Duration	5.681	0.176	4.094	6.579	1											
2. Recent Non-Concurrent Variety	1.038	0.573	0	1.792	0.017**	1										
3. Non-Concurrent Variety	5.115	1.352	0	7.483	0.118**	0.398**	1									
4. Recent Concurrent Variety	0.831	0.657	0	1.609	-0.045**	0.420**	0.195**	1								
5. Concurrent Variety	4.813	1.281	0	6.907	0.071**	0.367**	0.676**	0.266**	1							
6. Recent Non-Concurrent Variety x Non-Concurrent Variety	5.596	3.468	0	11.378	0.051**	0.819**	0.641**	0.635**	0.373**	1						
7. Recent Concurrent Variety x Concurrent Variety	4.199	3.565	0	9.676	-0.038*	0.421**	0.401**	0.434**	0.853**	0.477**	1					
8. Focal Experience	6.089	1.214	0	8.035	0.082**	0.357**	0.663**	0.204**	0.678**	0.637**	0.423**	1				
9. Team Size	1.617	0.130	0.693	2.079	0.117**	0.053**	0.078**	0.007	0.031*	0.088**	0.013	0.052**	1			
10. Average Individual Experience	5.662	1.036	0	7.312	0.110**	0.188**	0.641**	0.080**	0.663**	0.449**	0.281**	0.651**	0.071**	1		
11. Severe	0.043	0.203	0	1	0.036*	-0.029+	-0.074**	-0.010	-0.053**	-0.037**	-0.021	-0.064**	0.024	-0.048**	1	
12. Mild	0.291	0.454	0	1	-0.061**	0.011	-0.045*	0.026	-0.038+	-0.003	0.015	-0.038*	-0.024	-0.050**	-0.136**	1

+, * and ** denote significance at 10%, 5% and 1% levels respectively

Logged values of all variables except Severe and Mild

Table 4. Regression of Task Variety on Surgery Duration only for Lead Surgeons

Variable	Duration				
	Model: (1)	(2)	(3)	(4)	(5)
Non-Concurrent Variety		0.297**	0.317**	0.334**	0.325**
		(0.049)	(0.047)	(0.048)	(0.048)
Concurrent Variety			-0.301**	-0.360**	-0.339**
			(0.050)	(0.050)	(0.051)
Recent Non-Concurrent Variety				0.067**	0.066**
				(0.022)	(0.025)
Recent Concurrent Variety					-0.045*
					(0.022)
Non-Concurrent Variety x Recent Non-Concurrent Variety				0.032**	0.034**
				(0.011)	(0.013)
Concurrent Variety x Recent Concurrent Variety					-0.024*
					(0.011)
Focal Experience	-0.127**	-0.131**	-0.123**	-0.120**	-0.120**
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Team Size	0.109**	0.149**	0.132**	0.116**	0.115**
	(0.028)	(0.035)	(0.035)	(0.035)	(0.035)
Average Individual Experience	-0.031**	-0.030**	-0.022*	-0.023*	-0.021*
	(0.009)	(0.009)	(0.011)	(0.010)	(0.010)
Severe	0.025*	0.029*	0.027*	0.028*	0.028*
	(0.012)	(0.014)	(0.013)	(0.014)	(0.014)
Mild	-0.013*	-0.013*	-0.011+	-0.011+	-0.011+
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Constant	3.810**	3.796**	3.898**	3.968**	3.974**
	(0.110)	(0.112)	(0.112)	(0.114)	(0.115)
Observations (N)	3261	3261	3261	3261	3261
Adjusted R^2	0.143	0.147	0.160	0.162	0.163
Day of the Week Fixed Effect	Yes	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes	Yes	Yes	Yes

+, * and ** denote significance at 10%, 5% and 1% levels respectively

Table 5. Regression of Task Variety on Surgery Duration for all Surgeons

Variable	Duration				
	Model: (1)	(2)	(3)	(4)	(5)
Non-Concurrent Variety		0.171**	0.271**	0.269**	0.238**
		(0.019)	(0.023)	(0.023)	(0.024)
Concurrent Variety			-0.199**	-0.184**	-0.278**
			(0.025)	(0.025)	(0.025)
Recent Non-Concurrent Variety				0.033	0.050*
				(0.020)	(0.021)
Recent Concurrent Variety					-0.055*
					(0.023)
Non-Concurrent Variety x Recent Non-Concurrent Variety				0.014**	0.009*
				(0.004)	(0.004)
Concurrent Variety x Recent Concurrent Variety					-0.012**
					(0.004)
Focal Experience	-0.187**	-0.190**	-0.197**	-0.192**	-0.199**
	(0.017)	(0.020)	(0.023)	(0.023)	(0.023)
Team Size	0.067*	0.071**	0.070**	0.079**	0.081**
	(0.027)	(0.025)	(0.024)	(0.025)	(0.025)
Average Individual Experience	-0.031**	-0.018**	-0.019**	-0.019**	-0.019**
	(0.009)	(0.005)	(0.005)	(0.005)	(0.005)
Severe	0.032*	0.033*	0.030*	0.031*	0.031*
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Mild	-0.013*	-0.011+	-0.009	-0.008	-0.008
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Constant	4.110**	4.231**	4.215**	4.238**	4.231**
	(0.110)	(0.111)	(0.110)	(0.111)	(0.111)
Observations (N)	3261	3261	3261	3261	3261
Adjusted R^2	0.143	0.163	0.178	0.180	0.182
Day of the Week Fixed Effect	Yes	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes	Yes	Yes	Yes

+, * and ** denote significance at 10%, 5% and 1% levels respectively

Table 6. Regression of Task Variety on Surgery Duration with Instrumental Variables Approach

Variable	Non-Concurrent Variety	Duration	Non-Concurrent Variety	Duration
	Model: (1)	(2)	(3)	(4)
Non-Concurrent Variety		0.186**		0.201**
		(0.015)		(0.021)
Public Holidays	0.002**		0.002**	
	(0.000)		(0.000)	
Days of Vacations			0.001**	
			(0.000)	
Focal Experience	0.045**	-0.135**	0.047**	-0.098**
	(0.017)	(0.035)	(0.017)	(0.035)
Team Size	-3.057**	0.177**	-3.065**	0.176**
	(0.060)	(0.057)	(0.060)	(0.057)
Average Individual Experience	0.075*	-0.073**	0.077*	-0.078**
	(0.034)	(0.011)	(0.036)	(0.010)
Severe	0.104**	0.026*	0.104**	0.024*
	(0.034)	(0.013)	(0.034)	(0.012)
Mild	-0.004	-0.009	-0.005	-0.010
	(0.015)	(0.015)	(0.015)	(0.017)
Constant	5.518**	4.269**	5.532**	4.371**
	(0.123)	(0.764)	(0.133)	(0.726)
Observations (N)	3273	3261	3273	3261
Adjusted R^2	0.548	0.138	0.557	0.139
Day of the Week Fixed Effect	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes	Yes	Yes

+, * and ** denote significance at 10%, 5% and 1% levels respectively

Table 7. Regression of Task Variety on Surgery Duration with Instrumental Variables Approach

Variable	Concurrent Variety	Duration	Concurrent Variety	Duration
	Model: (1)	(2)	(3)	(4)
Concurrent Variety		-0.125**		-0.136**
		(0.027)		(0.032)
Public Holidays	0.001**		0.002**	
	(0.000)		(0.000)	
Days of Vacations			0.002**	
			(0.000)	
Focal Experience	0.047**	-0.282**	0.052**	-0.284**
	(0.018)	(0.073)	(0.018)	(0.075)
Team Size	-2.924**	0.169**	-2.906**	0.170**
	(0.058)	(0.049)	(0.058)	(0.049)
Average Individual Experience	0.045+	-0.073**	0.039*	-0.074**
	(0.024)	(0.016)	(0.019)	(0.016)
Severe	0.091**	0.027*	0.093**	0.028*
	(0.033)	(0.012)	(0.033)	(0.013)
Mild	-0.002	-0.011	-0.002	-0.011
	(0.015)	(0.020)	(0.015)	(0.016)
Constant	4.474**	5.659**	4.431**	5.049**
	(0.228)	(1.214)	(0.127)	(1.019)
Observations (N)	3273	3261	3273	3261
Adjusted R^2	0.548	0.133	0.550	0.133
Day of the Week Fixed Effect	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes	Yes	Yes

+, * and ** denote significance at 10%, 5% and 1% levels respectively

Table 8. Regression of Task Variety on Surgery Duration with Instrumental Variables Approach

Variable	Non-Concurrent Variety	Concurrent Variety	Duration
	Model: (1)	(2)	(3)
Non-Concurrent Variety			0.246**
			(0.018)
Concurrent Variety			-0.163**
			(0.017)
Public Holidays	0.002**	0.002**	
	(0.000)	(0.000)	
Days of Vacations	0.001**	0.002**	
	(0.000)	(0.000)	
Focal Experience	0.047**	0.052**	-0.107**
	(0.017)	(0.018)	(0.039)
Team Size	-3.065**	-2.906**	0.146**
	(0.060)	(0.058)	(0.044)
Average Individual Experience	0.077*	0.039*	-0.073**
	(0.036)	(0.019)	(0.017)
Severe	0.104**	0.093**	0.025*
	(0.034)	(0.033)	(0.012)
Mild	-0.005	-0.002	-0.009
	(0.015)	(0.015)	(0.007)
Constant	5.532**	4.431**	5.764**
	(0.133)	(0.127)	(0.712)
Observations (N)	3273	3273	3261
Adjusted R^2	0.557	0.550	0.153
Day of the Week Fixed Effect	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes	Yes

+, * and ** denote significance at 10%, 5% and 1% levels respectively