DePaul University

From the SelectedWorks of James A. Belohlav

October, 2008

Examining Quality Improvement Programs: The Case of Minnesota Hospitals

James A. Belohlav, *DePaul University*

Examining Quality Improvement Programs: The Case of Minnesota Hospitals

John R. Olson, James A. Belohlav, Lori S. Cook, and Julie M. Hays✠

Objective. To determine if there is a hierarchy of improvement program adoption by hospitals and outline that hierarchy.

Data Sources. Primary data were collected in the spring of 2007 via e-survey from 210 individuals representing 109 Minnesota hospitals. Secondary data from 2006 were assembled from the Leapfrog database.

Study Design. As part of a larger survey, respondents were given a list of improvement programs and asked to identify those programs that are used in their hospital.

Data Collection/Data Extraction. Rasch Model Analysis was used to assess whether a unidimensional construct exists that defines a hospital's ability to implement performance improvement programs. Linear regression analysis was used to assess the relationship of the Rasch ability scores with Leapfrog Safe Practices Scores to validate the research findings.

Principal Findings. The results of the study show that hospitals have widely varying abilities in implementing improvement programs. In addition, improvement programs present differing levels of difficulty for hospitals trying to implement them. Our findings also indicate that the ability to adopt improvement programs is important to the overall performance of hospitals.

Conclusions. There is a hierarchy of improvement programs in the health care context. A hospital's ability to successfully adopt improvement programs is a function of its existing capabilities. As a hospital's capability increases, the ability to successfully implement higher level programs also increases.

Key Words. Quality improvement, program effectiveness, program implementation, Rasch model analysis

Over $2 trillion was spent on health care in the United States in 2006, the highest level of per capita spending in the world. With health insurance premiums doubling every 5 years, it is predicted that a family's annual costs for health insurance will be $22,000 by the year 2010 (DoBias and Evans 2006).

Further, it is estimated that between 44,000 and 98,000 preventable deaths occur every year as a result of "errors" in the health care system and preventable health care-related injuries result in costs of between $17 and $29 billion annually (Corrigan, Kohn, and Donaldson 2000). Given the preceding scenario, many policy makers have begun to question the value that is being delivered by the U.S. health care system.

In an effort to strengthen their operating performance, leading health care organizations are instituting a variety of programs to reduce costs, increase safety, and improve clinical outcomes in an increasingly aggressive health care marketplace. Six Sigma, lean management, Studer programs, and the 100K Lives Campaign are some examples of programs currently being employed by health care organizations. However, it is not uncommon to see reports of both the successes and the failures for each of these programs. The difficult question confronting hospital administrators is, "Why do some programs fail and others succeed?"

CUMULATIVE CAPABILITIES

Like health care institutions, traditional business organizations are also being confronted with a variety of choices for improving their performance. Because of intense global competition, manufacturing organizations are especially interested in understanding how operating relationships lead to improved performance. Ferdows and DeMeyer (1990) proposed that successful manufacturing operations utilized a sequence of management efforts and resources that could be visualized as a sand cone that consisted of mutually dependent layers. The operational progression was described as: (1) initially constructing a foundation that produces *quality outcomes*, (2) followed by increasing *dependability* within the operating system, (3) then enhancing the *speed and flexibility* of the system, and (4) finally emphasizing lasting *cost improvements*.

Address correspondence to John R. Olson, Ph.D., Associate Professor, Department of Decision Sciences, Opus College of Business, University of St. Thomas, 1000 LaSalle Avenue, Minneapolis, MN 55403; e-mail: jrolson2@stthomas.edu. James A. Belohlav, Ph.D., Professor and Lori S. Cook, Ph.D., Associate Professor, are with the Department of Management, DePaul University, Chicago, IL. Julie M. Hays, Ph.D., Associate Professor, was with the Department of Decision Sciences, Opus College of Business, University of St. Thomas, Minneapolis, MN.

*Deceased.

Based upon the sand cone framework proposed by Ferdows and DeMeyer, Roth (1996a) formulated a theory describing the development of manufacturing organizational capabilities as a cumulative process, the Competitive Progression Theory (CPT). The relationship among the stages was described as one where the lower level stages facilitated the accomplishment of upper level stages. Specifically, Roth (1996b) states, "higher level organizational knowledge-based competencies are required to move efficiently between successive stages" (p. 38.13). Rosenzweig and Roth (2004) tested the efficacy of the CPT model using a sample of high-technology companies. Their results found evidence of the hypothesized progression of companies through the four stages and that the progression was related to organizational performance.

Butler and Leong (2000) used their observation that the "Accreditation Manual for Hospitals is consistent with the principles of operations strategy being discussed in the general operations management literature" (p. 228) as an impetus to examine competitive operating priorities within the hospital setting. Their study found that hospital operating priorities included quality, cost, flexibility, and delivery. In particular, they found interrelationships among these four areas, which imply that the concept of a cumulative capability development process may be relevant to the health care setting.

PERFORMANCE IMPROVEMENT PROGRAMS

Hospitals, similar to organizations in other economic sectors, have begun to systematically employ a variety of improvement programs. Some of the programs have been developed in the health care environment while others have been adapted from different economic sectors (Shortell et al. 1995). Table 1 displays the correspondence of the improvement programs being examined in this study to the four stages proposed in the CPT. Evans (2008) provides an overview of how the varied improvement programs in Table 1 relate to each other. Essentially, he views the programs at the lower stages of the CPT model as consisting of total quality projects (improving services and work processes, etc.). In contrast, those programs at the higher stages of the model are viewed as integrating, larger scale initiatives, which provides a crosscutting framework that includes multiple functions or perhaps even the whole organization. Revere and Black (2003) provide the following commentary, "Using the Six Sigma metrics, internal project comparisons facilitate resource allocation while external project comparisons allow for benchmarking. Thus, the appli-

Table 1: Correspondence of Improvement Programs to the Stages of the Competitive Progression Theory (CPT)

Stage Classification	Description	Program	References
Cost	Once all of the systems and processes are integrated, a certain level of control results. Lower costs come from "leveraging the entire enterprise and its network of suppliers, customers, partners, and facility locations" (Roth 1996a, p. 34). Even as changes become necessary, the hospital is able to maintain lower costs relative to other hospitals.	ISO Supply Chain Management (SCM) Lean Management Malcolm Baldrige Award Six Sigma	van den Heuvel et al. (2006), Martin (2006) Hayes and Pisano (1994) Hendricks and Singhal (2001) van den Heuvel et al. (2006), Germaine (2007)
Flexibility	"Quality and delivery pave the way for flexibility due to improvements in subsystem predictability and know how. Proverbial firewalls at the boundaries are destroyed in order to get closer to suppliers, partners, and especially customers. Flexibility requires an inordinate closeness to customers" (Roth 1996a, p. 34).	Voice of the Customer (VOC) Quality Function Deployment (QFD) External Supplier Awards Customer Relationship Management (CRM) Internal Quality Awards	Olhager and West (2002) McComb, Green, and Compton (2007) Hendricks and Singhal (2001) Javalgi et al. (2005) Samuelsson and Nilsson (2002), van der Wiele, Dale, and Williams (2000)
Dependability	"Fast and reliable delivery is possible only after a company can control its product quality. Delivery builds upon product quality by expanding the need to share technical information and process knowledge cross-functionally and along the supply chain" (Roth 1996a, p. 34).	Balanced Scorecard Studer Programs Cross-Functional Teams 5 Million Lives Campaign FOCUS PDSA Pay Bonus	Kaplan and Norton (2005) Studer (2004), Berger (2007) Marsh and Stock (2006), Rosenzweig and Roth (2004) Sandrick (2007) Brown and Hare (2002) Ross (2001), Young and Barrett (1997)
Quality conformance	The initial stage focuses on traditional quality management. "Quality management (QM) is broadly defined as a set of tools, action plans and principles aimed at reducing defects and improving processes . . . successful firms concentrate their early program targets first on achieving 'fitness for use' quality" (Rosenzweig and Roth 2004, p. 357).	Employee Recognition Programs Benchmarking PDCA Statistical Process Control (SPC) 100K Lives Campaign Employee Suggestion Programs	Ginther (2000) Blizzard (2004) Kondo (1993) Coleman et al. (2001) Wachter and Pronovost (2006) Silos (1999)

The references describe why each improvement program belongs in the respective stage classification level.

cation of Six Sigma makes TQM efforts more successful" (p. 377). While specifically discussing Six Sigma in the hospital setting, Revere and Black's observations virtually mirror those presented by Evans in identifying the interrelationship of higher level activities to those occurring at a lower level. They point out that once a hospital achieves the ability to implement higher level programs such as Six Sigma, lower level TQM tools become more effective because the hospital is able to identify those parts of the organization that can benefit most from quality improvement (QI) projects.

Research on competitive operating priorities from a cumulative development perspective provides a systematic approach for examining decisions in the hospital setting. The present study examines the experiences of Minnesota hospitals in implementing a variety of improvement programs. With this research, we provide a strategic perspective for the hospital decision making process. In addition, we seek to examine the impact of these operating decisions on performance. In order to address these issues, we consider two research questions.

Regarding the first research question, Butler and Leong (2000) note, "Researchers widely agree on four general operations competitive priorities: costs, quality, delivery, and flexibility" (p. 228). These operating priorities have been viewed from the perspective of cumulative manufacturing capabilities both at the theoretical level by Ferdows and DeMeyer (1991) and at the empirical level by Rosenzweig and Roth (2004) and Schroeder, Bates, and Junttila (2002). Hospitals, similar to manufacturing plants, are unit-level operating organizations. It is reasonable to suggest that hospitals also progress through a capability-building process. Thus, it useful to examine whether a hospital's capability-building process is similar to the experience of manufacturing organizations.

    1. Does a cumulative development process exist that influences the ability of hospitals trying to operationalize improvement programs?

While a variety of research attempts to assess the efficacy of particular programs or practices, perhaps the most relevant finding relates to the number of programs used. Relative to implementation and usage, Corbett and Whybark (2001) report, "The larger the number and intensity of the practices used, the greater the performance of the firm" (p. 978). This study along with the cumulative capabilities research would imply that hospitals would be able to employ greater numbers of programs as they develop increased capabilities. Further, those hospitals having a greater ability to implement programs

should demonstrate higher levels of performance as stated in the second research question.

> 2. Is there a relationship between hospital performance and its ability to implement improvement programs?

METHOD

This study investigates the usage pattern of improvement programs by member hospitals of the Minnesota Hospital Association (MHA). A two-stage analysis of the data was performed. The first stage identifies the capability of hospitals in carrying out improvement programs and defines the relative difficulty of the improvement programs themselves. The second stage of the analysis examines the relationship between the pattern of program usage and hospital performance outcomes.

Sample

An e-survey was sent by the MHA to 143 Minnesota hospitals during the period January to June 2007. The 674 selected survey recipients included Chief Executive Officers (CEO), Chief Medical Officers (CMO), Chief Financial Officers (CFO), Directors of Nursing (DON), and Quality Directors (QD). The MHA did not have contact information for all types of individuals at every hospital; at some hospitals one individual had several roles and some hospitals had several individuals with the same role. Three follow-up e-mails and a postcard reminder were sent to individuals who had not yet responded. Responses were received from 210 individuals within 109 hospitals, resulting in an effective hospital response rate of 76 percent. Multiple responses from an individual hospital were combined to create the hospital improvement program profile. Overall, 32 percent of the respondents were business administrators (CEO, CFO, or COO), 28 percent were medical administrators (CMO or DON), 30 percent were QD, and 10 percent were other managerial positions in the hospital. Characteristics of the responding hospitals are shown in Table 2.

Program Measures

For this study, data on program usage were collected as part of a larger survey. The survey respondents were given a list of programs and asked to select the

Table 2:   Respondent Demographics ($N = 109$ Hospitals)

	Mean	Median	Range
Revenue ($1,000,000)*	80	25	4–834
Case-mix index[†]	1.25	1.17	0.90–2.03
Number of beds (%)			
25 or less	41		
26–50 beds	22		
51–100 beds	15		
>100	22		
Hospital type (%)			
Large hospital	15		
Inpatient community hospital	71		
Not categorized	16		
Hospital location (%)			
Urban	38		
Rural	62		

*Data were available for 102 hospitals.
[†]Data were available for 57 hospitals.

programs implemented in their hospital. Table 1 presents the programs examined in this study, which include functional as well as larger scale organizational programs. The list of programs used in this study was generated through a literature search and consultation with practitioners and academics. The specific programs used were not intended to be an exhaustive list but, rather, a representative set of programs.

Outcome Measures

The second stage of the analysis examines the capability of the hospitals relative to the publicly available demographic and quality data from the Leapfrog Group. The Leapfrog Group is a consortium of major companies and other large private and public health care purchasers. In 2001, the Leapfrog Group created its initial survey to assess hospital performance based on practices that are proven to reduce medical mistakes. The Leapfrog Group rates a hospital's progress relative to the 27 safe practice areas that were identified by the National Quality Forum. Specifically, we compared the Leapfrog Safe Practices Score from the Leapfrog Hospital Quality and Safety Survey (Leapfrog Group 2007) with an ability score which was generated from our analysis.

Rasch Method

The data were first analyzed using Rasch Model Analysis (RMA), which is described more generally as a latent trait analysis. This type of analysis tries to

define an underlying factor that cannot be observed or measured directly from other observable variables. Consequently, the independent variable in this analysis is latent rather than an observed variable. Borsboom, Mellenbergh, and Japp (2003) comment, "although the model cannot be tested directly for any given item because the independent variable is latent, it can be tested indirectly through its implications for the joint probability distribution of the item responses for a number of items" (p. 205).

RMA consists of a family of models. The original model developed by Georg Rasch (1960/1980) was proposed for analyzing dichotomous data. The principles of the original model have been extended to rating scales (Andrich 1978; Wright and Masters 1982) and graded items (Masters 1982; Linacre 1994). This analysis will utilize the original model for dichotomous data.

In this study, we will consider whether hospitals can be characterized by their use of improvement programs. The Rasch model will examine whether the data correspond to a specific, basic underlying structure. It will accomplish this task by employing two basic parameters, difficulty (δ) and ability (β). The most common form of the model is represented as

$$p = \exp(\beta - \delta)/[1 + \exp(\beta - \delta)]$$

It is expected that the usage of individual improvement programs in hospitals provide varying degrees of challenge, which is referred to as the Rasch difficulty parameter (δ). In a similar fashion, it is expected that individual hospitals possess varying levels of capability in implementing improvement programs, which is referred to as the Rasch ability parameter (β). The δ parameters are a linear continuous measure of difficulty and the β parameters are a linear continuous measure of ability. RMA proposes that improvement program difficulty (δ) and a hospital's ability to achieve a particular performance improvement program (β) can be located on the same primary latent variable. Specifically, the Rasch difficulty parameter (δ) is used to locate these performance improvement programs along a continuum of program capability. Similarly, the Rasch ability parameter (β) *conjointly* locates an organization, based on its ability to achieve particular performance improvement programs, along the same underlying continuum.

In considering data, the overall approach that RMA employs is different than traditional assessment methods. Smith (1996) appropriately notes, "In descriptive statistical methodology, fit statistics are used to discover a model that fits the data well enough that the data could be considered to have been generated by the model. In Rasch analysis, the model is already chosen. The purpose of the fit statistics is to aid in measurement quality" (p. 516). Thus,

RMA tries to fit the data to a predetermined model rather than uncovering a model that fits the data.

RMA was used in this study because of several distinctive characteristics. First, Rasch analysis can be used to assess whether a unidimensional construct exists that defines a hospital's ability to implement improvement programs, which is the program capability. RMA focuses on how well "individual differences can be mapped on a single real number line" (Andrich 1988, p. 9). Second, all data are transformed into an interval scale as part of the analysis (Bond and Fox 2007). As a consequence, the dichotomous choice items used in the survey are converted into equal interval units using a logarithmic transformation (i.e., logits) which allows the interpretation of the differences among hospitals in their ability to implement improvement programs, thus facilitating the second stage of the analysis. Finally, RMA allows the same set of data to be used for both estimating and testing solutions (Rasch 1960/1980; Wright and Stone 1979; Wright and Masters 1982). A more detailed technical discussion of the Rasch model used in this study can be found in Bond and Fox (2007) and Andrich (1988).

RESULTS

The first stage of analysis was the development of the Rasch model. When a Rasch model is able to be constructed, it indicates that there is an underlying dimension that is common to all of the variables. Several indicators are utilized in assessing the overall fit of the data to Rasch model: model reliability, item quality using point measure correlations, and measure quality using infit and outfit statistics. *Winsteps* version 3.63.2 was used to analyze the data.

Using RMA, we are able to independently assess the reliabilities for hospitals and improvement programs. The reliability analyses reveal the extent to which the program usage items yield an internally consistent measure. Hospital reliability is $r = 0.83$ and program reliability is $r = 0.97$. These values indicate a high degree of fit; thus providing an initial indication that a unidimensional representation of the data exists.

Next, the relationship of individual improvement program variables to the overall model is examined. The point measure correlation of each item with the total score indicates how well a program predicts the total number of programs supported. For this particular measure, the sign of the correlation should be positive and is the relevant aspect of the analysis, rather than the magnitude of the correlation. As shown in Table 3, all signs are positive with the exception of ISO/TS-certified programs. The negative correlation indi-

Table 3: Improvement Programs Scores

	Measure (in Logits)*	Standard Error (SE)	r_{pm}	Infit	z-Score	Outfit	z-Score
Hospital program							
Quality Function Deployment (QFD)	3.20	0.47	0.45	0.90	− 0.20	0.34	− 0.40
External or Supplier Award	2.77	0.41	0.47	0.96	− 0.10	0.39	− 0.50
Malcolm Baldrige Award	2.62	0.40	0.33	1.39	1.40	0.83	0.10
Statistical Process Control (SPC)	1.97	0.34	0.48	0.98	0.00	0.88	0.00
Voice of the Customer (VOC)	1.80	0.32	0.58	0.77	− 1.20	0.40	− 1.10
Six Sigma	1.80	0.32	0.56	0.82	− 0.90	0.45	− 1.00
Internal Quality Program	1.04	0.28	0.49	1.08	0.60	0.95	0.00
Customer Relationship Management (CRM)	0.64	0.26	0.46	1.15	1.10	1.13	0.50
Supply Chain Management (SCM)	0.30	0.25	0.57	0.94	− 0.50	0.94	− 0.10
Lean Management	0.10	0.25	0.49	1.16	1.30	1.07	0.40
Pay Bonus Plans	− 0.06	0.24	0.54	1.00	0.10	0.94	− 0.20
Studer or Studerlike Program	− 0.30	0.24	0.46	1.21	1.80	1.19	0.90
FOCUS PDSA	− 0.43	0.24	0.52	1.10	0.90	1.05	0.30
5 Million Lives Campaign	− 0.58	0.24	0.61	0.88	− 1.10	0.79	− 0.90
PDCA	− 0.75	0.24	0.60	0.90	− 0.90	0.78	− 0.90
Balanced Scorecard	− 1.35	0.24	0.56	0.99	− 0.10	0.91	− 0.20
Cross-Functional Teams	− 1.43	0.24	0.56	0.97	− 0.20	1.01	0.10
100K Lives Campaign	− 2.16	0.27	0.61	0.80	− 1.60	0.64	− 0.80
Employee Suggestion System	− 2.24	0.27	0.41	1.27	1.90	1.36	0.90
Benchmarking	− 3.31	0.33	0.48	1.09	0.50	0.78	− 0.10
Employee Recognition Programs	− 3.61	0.37	0.52	0.91	− 0.30	0.51	− 0.30
Excluded programs							
ISO/TS Certification	5.03	1.03	− 0.02	1.14	0.50	5.63	1.80

*The more negative a number is, the more often that an improvement program is accomplished. The more positive a number is, the less often that an improvement program is accomplished.

cates that the ISO/TS-certified program variable does not fit the overall model for hospitals and this variable was removed from further analysis.
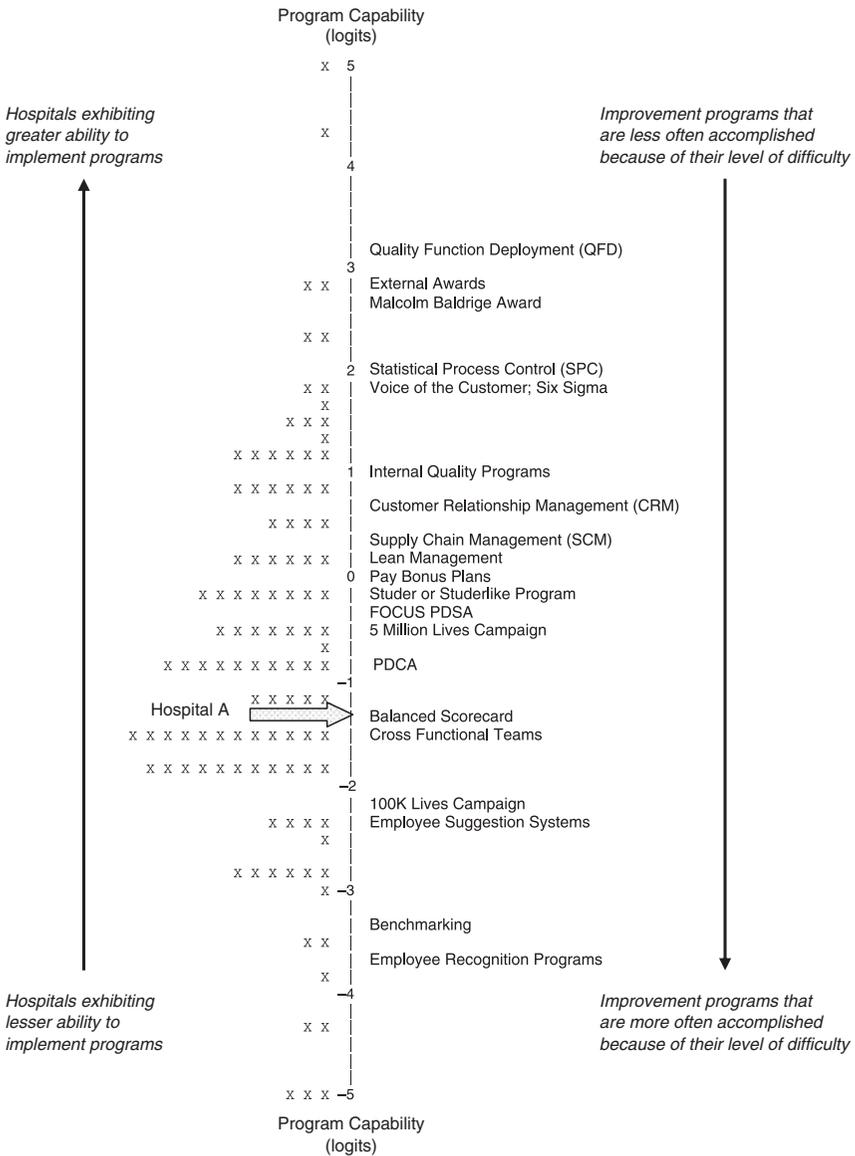
Finally, how well the individual items fit the model is considered. Mean square fit statistics are used to assess whether observations are in agreement with the Rasch model values. Two types of fit can be used in the Rasch model, outfit and infit statistics. The outfit measure is sensitive to unexpected observations of hospitals and programs that are relatively far from their position. The infit measure is sensitive to unexpected patterns of observations by hospitals and programs that are close to their position (Wright and Masters 1982). One way of examining individual item fit is by means of z-scores. The standardized z-scores for the fit statistics should fall within a -2 to $+2$ range for acceptable fit. As shown in Table 3, all of the remaining improvement programs are within the -2 to $+2$ standardized z-score range defining acceptable fit. Based on these results, we conclude that the data provide a good fit to the Rasch model, thus revealing the existence of a unidimensional latent trait that can be described as the program capability of hospitals.

Even though a Rasch model was able to be effectively constructed based on improvement program usage, its correspondence to the stage model presented in Table 1 also needs to be assessed. The measure scores in Table 3 report the difficulty of each program, the scores are ordered according to level of difficulty with the easier programs starting at the bottom. Assuming that all lower stage programs are implemented before the next higher level stage programs from Table 1, the Wilcoxon test, a nonparametric version of the t-test, was used to analyze whether a difference exists between the results of the Rasch model and the four-stage model presented in Table 1. No statistically significant differences were observed ($W = -6$, $n_{s/r} = 8$) between the improvement programs in the proposed four-stage model and the results presented by the Rasch model. Although some differences were present, the results provided by the Rasch model appeared to be consistent with the four-stage view of cumulative development.

The end result of the Rasch analysis is that all improvement programs receive a difficulty score and all hospitals receive an ability score. Both the difficulty score and the ability score are reported as logits. The results for both program location and hospital location are presented graphically in Figure 1, the variable map. The vertical axis in the center of the variable map serves as a ruler, with logits as a common unit of measurement, allowing the comparison of both hospitals and improvement programs.

On the right side of the map, the distribution of the improvement programs is shown. Those improvement programs that are located higher on the

Figure 1:   Variable Map

Program Capability
(logits)

Hospitals exhibiting greater ability to implement programs

Improvement programs that are less often accomplished because of their level of difficulty

Quality Function Deployment (QFD)

External Awards
Malcolm Baldrige Award

Statistical Process Control (SPC)
Voice of the Customer; Six Sigma

Internal Quality Programs

Customer Relationship Management (CRM)

Supply Chain Management (SCM)
Lean Management
Pay Bonus Plans
Studer or Studerlike Program
FOCUS PDSA
5 Million Lives Campaign

PDCA

Hospital A
Balanced Scorecard
Cross Functional Teams

100K Lives Campaign
Employee Suggestion Systems

Benchmarking

Employee Recognition Programs

Hospitals exhibiting lesser ability to implement programs

Improvement programs that are more often accomplished because of their level of difficulty

Program Capability
(logits)

map are accomplished less frequently than those that are lower on the map. Rasch modeling literature refers to this aspect as a program's *difficulty*. Within the health care context, two significant sources of difficulty arise because: (1) an improvement program can be difficult to deploy throughout a hospital or (2) a hospital is unable to understand how a particular improvement program can be effectively employed. The hierarchy from the easiest to the most difficult programs to implement exhibits a range that spans more than 6 logits. This result implies that a well-defined hierarchical structure exists among the improvement programs. Based upon the Rasch difficulty scores, there is a statistically significant difference in difficulty ($\sim z = 3.85$) between the bottom 25 percent of the programs and the top 25 percent of the programs, which approximates the first stage (quality conformance) and the highest level stage (low cost) presented in Table 1.

   The distribution of hospitals is located on the left side of the map. The hospitals in this study had widely varying abilities in carrying out programs. Hospitals located higher on the left side of the map have been able to adopt more programs than hospitals that are located lower on the map. Rasch modeling literature refers to this aspect as the hospital's *ability* to carry out a program. Based upon the Rasch ability scores, there is a statistically significant difference ($\sim z = 2.13$) between the bottom 10 percent of the hospitals and the top 10 percent of the hospitals in their ability to carry out programs. That is, the hospitals with the most ability were able to implement more programs than the hospitals with lesser levels of ability.

   What does the preceding statistical analysis tells us? First, the ordering of improvement programs that we obtained in this study is consistent with the stages presented in the Competitive Progression Theory. As a result, we would expect that the improvement programs representing the highest stage activities would be accomplished less often than improvement programs representing the lowest stage activities in this model (because they are more difficult), which we found to be true. Finally, because of the range of difficulty presented by the improvement programs, we would then expect that hospitals would have varying levels of capability in carrying out improvement programs. Indeed, the most capable hospitals were able to implement more programs than the least capable hospitals. Aside from the theoretical framework that proposes general reasons for the success or failure in implementing improvement programs, the practical significance of this analysis lies in the interpretation of the variable map for individual hospitals.

   The variable map is interpreted in terms of both hospitals and improvement programs. A program that is 1 logit higher than a hospital's position

means that it is twice as difficult to accomplish as a program that is at its level. The converse is also true. A program that is 1 logit lower than a hospital's position means that it is twice as easy to accomplish as a program that is at its level. The difficulty of adopting the program can also be described in terms of its probability of success. Understanding these relationships can guide an organization in making more effective operating decisions and more efficient resource allocations.

For example, if Hospital A as denoted by the arrow in Figure 1 (at the level of the Balanced Scorecard) decided to pursue FOCUS PDSA, there is a 27 percent probability that it will be able to successfully implement the program. If Hospital A decides to pursue the Malcolm Baldrige Award, the chances of success are even lower, <5 percent. However, if the hospital did not have an employee suggestion system and decided to implement that type of program the probability of successfully implementing that program is quite high, nearly 75 percent.

The probabilities of success cited in the previous paragraph are related to the difficulty of achieving a particular program. Probability of success like difficulty is determined by distances on the vertical ruler (Bond and Fox 2007). If a program that a hospital wants to accomplish is 1 logit higher than where the hospital is currently situated in terms of its ability, there is only a 27 percent probability that the program will be accomplished. If a program that a hospital wants to accomplish is 2 logits higher, it is three times as difficult and there is only a 12 percent probability that the program will be accomplished. If a practice that a hospital wants to accomplish is 3 logits higher, it is four times as difficult and there is only a 5 percent probability that the program will be accomplished. One can see that the farther away a program is from the ability of the hospital the greater the chance for failure.

It is important to note that these are only probabilities. Any selection of a program will generally require devoting a substantial amount of financial, human, and time resources in order to ensure that the program is successful. It should be noted that low probabilities do not preclude a hospital from attempting to implement more difficult programs. It simply means that greater effort and more resources will need to be expended by a hospital to ensure the successful implementation of the program.

The second stage of analysis views the overall impact of a hospital's ability to implement programs within its organization relative to its operating performance. To measure the impact on performance, we used the results of the Rasch analysis in conjunction with the Leapfrog score. Because inclusion in the Leapfrog database is voluntary, it should be noted that only 48 of the

hospitals taking part in this study also participated in the Leapfrog survey, which is 44 percent of the responding hospitals. A regression model was built using the Rasch hospital ability score (independent variable) and its respective Leapfrog score. Initially, we included case-mix index, number of beds, and urban or rural location in the model to control for differences among hospitals. However, none of these control variables were found to be significant and they were removed from the model. The results from the final model indicate the Rasch ability score is significantly correlated with the Leapfrog score, $F(1.47) = 16.846$, $p < .01$. The $r^2 = 0.264$ indicates that over 25 percent of the total variance in the Leapfrog score is explained by the Rasch model ability score. That is, the hospitals with greater numbers of implemented programs tended to have higher scores on the Leapfrog Hospital Quality and Safety Survey.

DISCUSSION

Implications

As hospital administrators sort through the various improvement programs at their disposal, the obvious question is: "Where do we start?" There are several key points arising from this research that can assist in the decision making process. First, the highest level programs are statistically more difficult to accomplish than the lower level programs. While there were some differences in ordering of improvement programs among the stages proposed in Table 1 and the actual results, overall the two were not found to be statistically different. That is, the research found the hierarchy of adoption of improvement programs among Minnesota hospitals to be consistent with the four-stage development model. Second, the most capable hospitals were able to implement more programs than the least capable hospitals. More importantly, the hospital capability, as identified by the Rasch model, resulted in a statistically significant correlation with the Leapfrog Hospital Quality and Safety Survey indicators.

The current study can provide some clarification relative to existing research. For example, a recent study by Weiner et al. (2005) reports that some of their findings were contrary to their initial expectations. In particular, greater involvement in QI efforts by multiple hospital units resulted in lower values on quality indicators. They offer several interpretations of their findings. One of these interpretations was, "extensive involvement by multiple units could yield little to no improvement in hospital-level quality indicators if

QI projects work at cross purposes because of poor coordination or inappropriate sequencing" (p. 326). Our findings tend to support the preceding interpretation of Weiner et al.'s apparently contradictory results. If a hospital implements elements of improvement programs in the wrong sequence, it would most likely face an increase in costs or in the worst-case experience outright failure. For instance, trying to achieve success in a Six Sigma program would be substantially more difficult if many of the lower level initiatives were not already implemented.

Limitations and Future Research

Several limitations exist in this study. First, the sample used to develop the variable map was based exclusively on hospitals located in Minnesota. Generally, Minnesota hospitals operate as not-for-profit organizations; therefore, the findings may not be applicable to all types of hospitals. Second, we assumed that the type of respondents that we selected, higher managerial levels, would have knowledge of the various programs implemented within their organizations. Even though the respondents occupied higher level positions, they may not have been familiar with activities in all areas of the hospital. Third, not all of the hospitals taking part in this study were contained in the Leapfrog database. As a result, the statistical relationship found between hospital ability and Leapfrog Hospital Quality and Safety Survey could be overstated. Finally, the improvement programs examined in this study consisted of a representative list, thus some noteworthy programs may have been excluded.

As hospitals emphasize a more systematic approach to improving the quality of care, patient safety, and business processes, future research should examine whether organizational efforts focusing on only one type of improvement program is sufficient. For example, Benitez et al. (2007) indicate that Six Sigma and Quality Function Deployment have been successfully utilized together to create and maintain effective hospital processes. Caldwell, Brexler, and Gillem (2005) advocate the adoption of a combination of lean management and Six Sigma practices to overcome existing deficiencies. It would be useful to understand whether the complexities of combinations of programs are more difficult to implement.

CONCLUSIONS

A hospital's ability to successfully adopt improvement programs is hierarchical in nature, some programs are easier to implement than others. Imple-

menting higher level programs becomes easier when lower level programs have already been implemented within the organization. By understanding its own level of capability, a hospital can effectively identify its stage of development thus allowing scarce resources to be deployed more efficiently. This research can provide a roadmap for the selection of improvement programs so that the probability of successful implementation is maximized.

ACKNOWLEDGMENTS

REFERENCES

Andrich, D. 1978. "A Rating Scale Formulation for Ordered Response Categories." *Psychometrika* 43 (4): 561–73.
———. 1988. *Rasch Models for Measurement*. Newbury Park, CA: Sage Publications.
Benitez, Y., L. Forrester, C. Hurst, and D. Turpin. 2007. "Hospital Reduces Medication Errors Using DMAIC and QFD." *Quality Progress* 40 (1): 38–45.
Berger, S. H. 2007. "Treating Technology as a Luxury?: 10 Necessary Tools." *Healthcare Financial Management* 61 (2): 63–9.
Blizzard, R. 2004. "Patient Satisfaction Only Seven Steps Away?" *Gallup Poll Tuesday Briefing*: 1–3: June 22, 2004.
Bond, T. G., and C. M. Fox. 2007. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 2nd Edition. London: Lawrence Erlbaum Associates.
Borsboom, D., G. J. Mellenbergh, and V. H. Japp. 2003. "The Theoretical Status of Latent Variables." *Psychological Review* 110 (2): 203–19.
Brown, P., and D. Hare. 2003. "Rapid Cycle Improvement: Controlling Change." *Journal of the Arkansas Medical Society* 98 (10): 2–3.
Butler, T. W., and G. K. Leong. 2000. "The Impact of Operations Competitive Priorities on Hospital Performance." *Health Care Management Science* 3 (3): 227–35.

Caldwell, C., J. Brexler, and T. Gillem. 2005. "Engaging Physicians in Lean Six Sigma." *Quality Progress* 38 (11): 42–7.

Coleman, S. Y., G. Arunakumar, F. Foldvary, and R. Feltham. 2001. "SPC as a Tool for Creating a Successful Business Measurement Framework." *Journal of Applied Statistics* 28 (3): 325–34.

Corbett, L. M., and D. C. Whybark. 2001. "Searching for the Sandcone in the GMRG Data." *International Journal of Operations and Production Management* 21 (7): 965–80.

Corrigan, J., L. T. Kohn, and M. S. Donaldson (eds) 2000. *To Err Is Human: Building a Safer Health System.* Washington, DC: National Academies Press.

DoBias, M., and M. Evans. 2006. "Mixed Signals—The CMS 10-Year Spending Projections Inspire Both Hope and Skepticism, and Leave Plenty of Room for Lobbyists." *Modern Healthcare* 36 (9): 6–8.

Evans, J. R. 2008. *Quality and Performance Excellence: Management, Organization and Strategy.* 5th Edition. Mason, OH: South-Western Cengage Learning.

Ferdows, K., and A. DeMeyer. 1990. "Lasting Improvements in Manufacturing Performance: In Search of a New Theory." *Journal of Operations Management* 9 (2): 168–84.

Germaine, J. 2007. "Six Sigma Plan Delivers Stellar Results." *Materials Management in Health Care* 16 (4): 20–6.

Ginther, C. 2000. "Incentive Programs That Really Work." *HR Magazine* 45 (8): 117–20.

Hayes, R. H., and G. P. Pisano. 1994. "Beyond World-Class: The New Manufacturing Strategy." *Harvard Business Review* 72 (1): 77–87.

Hendricks, K. B., and V. R. Singhal. 2001. "Firm Characteristics, Total Quality Management, and Financial Performance." *Journal of Operations Management* 19 (3): 269–85.

Javalgi, R. G., T. W. Whipple, A. K. Ghosh, and R. B. Young. 2005. "Market Orientation, Strategic Flexibility, and Performance: Implications for Services Providers." *Journal of Services Marketing* 19 (4): 212–21.

Kaplan, R. S., and D. R. Norton. 2005. "The Balanced Scorecard: Measures That Drive Performance." *Harvard Business Review* 83 (7): 172–80.

Kondo, Y. 1993. "Quality Education in Japan." *Total Quality Management* 4 (2): 115–25.

Leapfrog Group. 2007. "Leapfrog Hospital Quality and Safety Survey Results" [accessed on August 16, 2007]. Available at http://www.leapfroggroup.org/cp

Linacre, J. M. 1994. *Many-Facet Rasch Measurement.* Chicago: MESA Press.

Marsh, S. J., and G. N. Stock. 2006. "Creating Dynamic Capability: The Role of Intertemporal Integration, Knowledge Retention, and Interpretation." *Journal of Product Innovation Management* 23 (5): 422–36.

Martin, B. 2006. "A Blueprint to Build a Cost-Effective Supply Chain." *Materials Management in Health Care* 15 (9): 32–4.

Masters, G. N. 1982. "A Rasch Model for Partial Credit Scoring." *Psychometrika* 47 (2): 149–74.

McComb, S. A., S. G. Green, and W. D. Compton. 2007. "Team Flexibility's Relationship to Staffing and Performance in Complex Projects: An Empirical Analysis." *Journal of Engineering and Technology Management* 24 (4): 293–313.

Olhager, J., and B. M. West. 2002. "The House of Flexibility: Using the QFD Approach to Deploy Manufacturing Flexibility." *International Journal of Operations and Production Management* 22 (1): 50–79.

Rasch, G. 1960/1980. *Probabilistic Models for Intelligence and Attainment Tests.* Chicago: University of Chicago Press.

Revere, L., and K. Black. 2003. "Integrating Six Sigma with Total Quality Management: A Case Example for Measuring Medication Errors." *Journal of Healthcare Management* 48 (6): 377–91.

Rosenzweig, E. D., and A. V. Roth. 2004. "Towards a Theory of Competitive Progression: Evidence from High-Tech Manufacturing." *Production and Operations Management* 13 (4): 354–68.

Ross, J. P. 2001. "Keeping the Focus on Satisfaction." *Healthcare Executive* 16 (6): 62–3.

Roth, A. V. 1996a. "Achieving Strategic Agility through Economies of Knowledge." *Strategy and Leadership* 24 (2): 30–7.

———. 1996b. "Neo-Operations Strategy." In *Handbook of Technology Management*, edited by G. Gaynor, pp. 38.1–38.44. New York: McGraw Hill.

Samuelsson, P., and L. Nilsson. 2002. "Self-Assessment Practices in Large Organisations: Experiences from Using the EFQM Excellence Model." *International Journal of Quality and Reliability Management* 19 (1): 10–23.

Sandrick, K. 2007. "The Importance of Getting Boards on Board." *Trustee* 60 (10): 16.

Schroeder, R. G., K. A. Bates, and M. A. Junttila. 2002. "A Resource-Based View of Manufacturing Strategy and the Relationship to Manufacturing Performance." *Strategic Management Journal* 23 (2): 105–17.

Shortell, S. M., J. L. O'Brien, J. M. Carman, R. W. Foster, E. F. X. Hughes, H. Boerstler, and E. J. O'Connor. 1995. "Assessing the Impact of Continuous Quality Improvement/Total Quality Management: Concept versus Implementation." *Health Services Research* 30 (2): 377–401.

Silos, I. M. 1999. "Employee Involvement—A Component of Total Quality Management." *Production and Inventory Management Journal* 40 (1): 56–65.

Smith, R. M. 1996. "Polytomous Mean-Square Fit Statistics." *Rasch Measurement Transactions* 10 (3): 516–7.

Studer, Q. 2004. "Healthcare Service: Moving from Good to Great." *Marketing Health Services* 24 (2): 20–5.

van den Heuvel, J., A. J. Bogers, R. J. Does, S. L. van Dijk, and M. Berg. 2006. "Quality Management: Does It Pay Off?" *Quality Management in Health Care* 15 (3): 137–49.

van der Wiele, T., B. Dale, and R. Williams. 2000. "Business Improvement through Quality Management Systems." *Management Decision* 38 (1/2): 19–23.

Wachter, R. M., and P. J. Pronovost. 2006. "The 100,000 Lives Campaign: A Scientific and Policy Review." *Joint Commission Journal on Quality and Patient Safety* 32 (11): 621–7.

Weiner, B. J., J. A. Alexander, S. M. Shortell, L. C. Baker, M. Becker, and J. J. Geppert. 2005. "Quality Improvement Implementation and Hospital Performance on Quality Indicators." *Health Services Research* 41 (2): 307–34.

Wright, B. D., and G. Masters. 1982. *Rating Scale Analysis.* Chicago: Mesa Press.

Wright, B. D., and M. H. Stone. 1979. *Best Test Design.* Chicago: Mesa Press.
Young, D. W., and D. Barrett. 1997. "Managing Clinical Integration in Integrated Delivery Systems: A Framework for Action." *Hospital and Health Services Administration* 42 (2): 255–79.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix SA1. Author Matrix.

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