

7-2015

Creating Quality Metrics for the Capital Facilities Delivery Industry

Kelsey D. Lamb

University of Arkansas, Fayetteville

Follow this and additional works at: <http://scholarworks.uark.edu/etd>



Part of the [Industrial Engineering Commons](#), and the [Industrial Technology Commons](#)

Recommended Citation

Lamb, Kelsey D., "Creating Quality Metrics for the Capital Facilities Delivery Industry" (2015). *Theses and Dissertations*. 1201.
<http://scholarworks.uark.edu/etd/1201>

This Thesis is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, ccmiddle@uark.edu.

Creating Quality Metrics for the Capital Facilities Delivery Industry

Creating Quality Metrics for the Capital Facilities Delivery Industry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Industrial Engineering

by

Kelsey Lamb
University of Arkansas
Bachelor of Science in Industrial Engineering, 2013

July 2015
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Dr. Kim LaScola Needy
Thesis Director

Dr. Robert Ries
Committee Member

Dr. Shengfan Zhang
Committee Member

Abstract

A standard for industry-wide quality metrics, similar to safety metrics, for the capital facilities delivery industry is investigated and developed with the goal to motivate industry-wide adoption of this new quality metric program. A survey of the literature found key differences between capital facilities and other industries, mainly that other industries focus on the quality of the product while capital facilities focus on maintaining the cost and schedule of a project. Creating a framework that can be used across the industry can potentially lead to uniform improvement of quality for the entire industry. Quality is defined for the capital facilities delivery industry and a quality pyramid that categorizes both leading and lagging indicators and can be used across the project lifecycle is presented. Lagging indicators are developed as set of quality events that were used to create a quality metric. A group of subject matter experts (SME), from various sectors of the industry, were assembled to aid in the research effort and validate results. Pilot data was collected via SME companies to test hypotheses and to better understand the industries' quality performance. Data is normalized and a quality performance rate (QPR) metric is developed. Statistical analysis of this pilot data shows differences amongst respondents. Future work is described towards the goal of creating an industry-wide quality metric.

Table of Contents

1. Introduction	1
1.1. Background Information	1
1.2. Motivation for the Study	1
1.3. Problem Statement	2
1.4. Report Layout and Limitations	3
2. Literature Review	5
2.1. Introduction	5
2.2. Body of Knowledge in Improving Quality and Developing Quality Metrics.....	5
2.2.1. Normalization	9
2.3. Quality metrics outside the capital facilities delivery industry	10
2.3.1. Healthcare	10
2.3.2. Pharmaceutical	11
2.3.3. Department of Energy	16
2.3.4. Safety Metrics	18
2.4. Quality Metrics for the Capital Facilities Delivery Industry.....	19
2.4.1. Quality Metrics in the US	19
2.4.2. Quality Metrics in other Countries	20
2.5. Conclusion	22
3. Methodology.....	23
3.1. Introduction	23
3.2. Quality Framework	25
3.2.1. Quality Definition	25
3.2.2. Quality Pyramid	26
3.2.3. Data Collection Tool.....	27
3.3. Data Analysis	29
3.3.1. Normalization	30
3.3.2. Hypotheses	30
3.4. Methodology Conclusion.....	33
4. Results	34
4.1. Introduction	34

4.2.	Quality Framework	34
4.2.1.	Quality Definition	34
4.2.2.	Quality Pyramid	34
4.3.	Data Collection Tool	36
4.4.	Data Analysis	37
4.4.1.	Demographic Data	37
4.4.2.	Metrics Data	42
4.4.3.	Hypotheses	46
4.4.4.	Data Analysis Conclusion	52
5.	Conclusions and Future Work	54
5.1.	Summary of research goals and objectives	54
5.2.	Summary of Research Outcomes	55
5.2.1.	Outcomes of the Literature Review	55
5.2.2.	Observations from the Data Collection	55
5.2.3.	Outcomes of the Data Analysis	58
5.3.	Conclusions	59
5.4.	Future Work	59
	References	62
	Appendix A: Illustration Document	65
	Appendix B: Confidentiality Agreement	72
	Appendix C: Survey Instructions	74
	Appendix D: Qualtrics Survey	76
	Appendix E: IRB Letter	78

List of Tables

Table 2-1. Quality Definitions	5
Table 2-2. Pharmaceutical Quality Metrics (Mendivil et al, p. 544-545)	13
Table 2-3. Performance Metric Programs from Various Countries	21
Table 2-4. Quality Metrics summarized from Cha (2011)	22
Table 4-1: Quality Event Definitions	35
Table 4-2. Number of Submitted Projects	38
Table 4-3. Number of Submitted Projects	38
Table 4-4. Number of Submitted Projects per Project Size	38
Table 4-5. Number of Submitted Projects per Industry Type	39
Table 4-6. Number of Projects per Prevention Activity	40
Table 4-7. Self-Reported Perception of Quality	41
Table 4-8. Number of Activities/Events by Type and per Project	42
Table 4-9. Owner Descriptive Statistics for Activities/Events per Project	43
Table 4-10. Contractor Descriptive Statistics for Activities/Events per Project	43
Table 4-11. Number of Normalized Activities/Events by Type	44
Table 4-12. Owners Descriptive Statistics for Normalized Activities/Events per Labor Hour	46
Table 4-13. Contractors Descriptive Statistics for Normalized Activities/Events per Labor Hour	46
Table 4-14. Hypothesis Statements	47
Table 4-15. Hypothesis 1 Correlation Analysis	48
Table 4-16. Hypothesis 1 Subgroups	48
Table 4-17. Hypothesis 2 Correlation Analysis	49
Table 4-18. Hypothesis 2 with Subgroups	50
Table 4-19. Hypothesis 3 Correlation Analysis	50
Table 4-20. Average Proportions of Events	52
Table 4-21. Hypothesis 5 Owners Analysis	52
Table 4-22. Hypothesis 5 Contractors Analysis	52

Table of Figures

Figure 2-1. Heinrich's Safety Pyramid (Heinrich, 1959)	18
Figure 2-2. Safety Pyramid from ConocoPhillips Marine (Masimore, p.11)	19
Figure 3-1. Example of the OSHA Safety Pyramid	26
Figure 4-1. Quality Pyramid	35
Figure 4-2. Percentage of Prevention Activities used per Project	40
Figure 4-3. Self-Reported Perceptions of Quality Graph	41
Figure 4-4. Average Normalized Events/200K Labor Hours Company Comparison	45
Figure 4-5. Descending Normalized Events/200K Labor hours per project	45
Figure 4-6. Hypothesis 2 Graph	49

1. Introduction

1.1. Background Information

In the capital facilities delivery industry, the two main factors that are used to measure the success of a project are cost and schedule. Customers, or owner companies, primarily want the project to be done within budget and on time. In recent years, the industry has seen a push to execute a project within a high quality standard as well. Companies that consistently produce high quality products and services will have high customer loyalty and a strong corporate brand. For this reason, many companies use quality metrics to improve their business systems. However, companies have gone about this endeavor independently from the industry which has led to many varying definitions of quality, measures of quality, and uses of quality metrics.

The research presented for this thesis is part of an ongoing effort to create quality metrics for the capital facilities delivery industry. This effort is funded by the Construction Industry Institute (CII), a research organization dedicated to the advancement of knowledge and practices within the industry. Currently, CII has more than 130 leading owner, engineering-contractor, and supplier companies enrolled as members of the institution. These members participate as subject matter experts (SMEs) in research projects that are of interest to them. The research team for “Creating Quality Metrics for the Capital Facilities Delivery Industry,” called RT313, is comprised of 24 SMEs from both owners and contractor companies and academics. The academics are from the University of Arkansas and the University of Florida.

1.2. Motivation for the Study

A review of the literature indicates that the only metrics that the capital facilities delivery industry collects uniformly are safety metrics mandated by the Occupational Safety and Health Administration (OSHA) and efficiencies of project costs and schedule. There is no existing quality management system (QMS) specifically for the industry that allows any and all sectors to

collect quality metrics on the actual deliverable. What the industry really needs is a way to assess quality performance and drive improvement. This research provides a set of quality metrics that can be uniformly applied across the industry to assess the quality performance of capital facilities for the first time.

A further review of the literature finds that other industries are discovering a need for a QMS such as this as well. In particular, the Food and Drug Administration (FDA) is working on a parallel study to develop quality metrics for the pharmaceutical industry. While the organization's work is also ongoing, a recent update during a conference suggests that the FDA is moving toward a set of metrics that include a severity scale and the ability to compare across both projects and industries. They have also indicated that the industry would like to be able to use this metrics program for contract acquisitions and to develop a culture of quality that will lead to improvement rather than constantly fixing issues without dealing with the root cause. All of these findings and realizations mirror those of the capital facilities delivery industry and help to validate CII's motivation for the study.

1.3. Problem Statement

The essential question for this research is: Can a standard method and set of standard metrics be established that can be used to effectively measure, categorize, and benchmark quality performance across the project delivery process? Specifically, this research has a primary objective to develop a set of quality metrics that allows both owners and contractors within the capital facilities delivery industry to:

- Replicate as much as possible the characteristics present in today's accepted safety metrics;
- Define what a quality failure is;

- Identify areas for improvement;
- Establish a common understanding (common language relative to quality similar to the common language utilized in measuring safety performance);
- Benchmark current performance and identify trends both internally and across the industry;
- Support ability to initiate pro-active actions to assure maintaining or improving quality in lieu of continually reacting to quality issues;
- Identify best practices and provide input that allows for work process improvements that demonstrate value to the business by positively impacting the “bottom line” (quantifying the cost of failure to support maintaining quality and understanding the impact of non-conformance); and
- Demonstrate quality performance relative to the industry to internal and external customers and use the performance indicated by the metrics to differentiate themselves.

A specific secondary objective of this research is to support a culture change where quality issues are identified, reported and measured similar to the changes that occurred in awareness of safety issues resulting in the current safety conscious culture that encourages and rewards continuing improvement (change behaviors).

1.4. Report Layout and Limitations

As previously mentioned, the research detailed in this thesis is based on an ongoing research project being funded by CII. As a result, this study is a subset of a large project and is not meant to have final conclusions drawn from its findings. Furthermore, this study is currently limited to developing lagging indicators for quality performance of capital projects. Leading indicators will be investigated and developed in further studies.

Chapter 2 of this thesis will present a review of the literature including a look at the general body of knowledge (BOK) relating to quality and quality metrics, and quality metrics both inside and outside of the capital facilities delivery industry. Next, Chapter 3 describes the approach and methodology that was used to develop the deliverables of this research study as well as the tools used for data collection and data analysis. Chapter 4 presents the results from the pilot data collection efforts including the developed set of quality metrics and frameworks and the outcomes of the data analysis and Chapter 5 details the conclusions from the pilot data collection and discusses future work.

2. Literature Review

2.1. Introduction

The literature review for creating quality metrics in the capital facilities delivery industry focused on studies that reported on the use of quality metrics within an organization or industry. Information about quality and quality metrics currently being used were collected based on three different categories of interest:

- Documents focusing on building the body of knowledge in improving quality and developing quality metrics;
- Quality metric practices currently being used in other industries;
- Quality metric practices generally used country-wide (or within a specific country) for the capital facilities delivery industry in the U.S. and other countries around the world.

2.2. Body of Knowledge in Improving Quality and Developing Quality Metrics

With the goal being to create new metrics for the capital facilities delivery industry, the first step was to obtain a clear understanding of the quality process and define key terms and elements relative to the industry. Many recognized quality leaders and professional societies have developed their own definitions of quality as depicted in Table 2-1.

Table 2-1. Quality Definitions

Individual/Society Name	Definition of Quality
Juran (1980)	Fitness of use.
Deming (1986)	(a) A product that conforms to a set of standards, and (b) a product that meets consumer wants and needs.
American Society for Quality (2013)	A subjective term for which each person has his or her own definition. In technical usage, quality can have two meanings: (1) the characteristics of a product or service that bear on its ability to <i>satisfy</i> stated or implied needs and (2) a product or service <i>free of deficiencies</i> .
Crosby (1962)	Conformance to requirements.
Taguchi (1986)	A product or service has good quality if it performs its intended functions without variability, and causes little loss through harmful side effects, including the cost of using it.

These definitions were used as a reference during a face-to-face meeting of RT313 that resulted in the development of a definition of quality as it pertains to the capital facilities delivery industry.

Once a working definition of quality was established, efforts were focused on performance assessment. Thor (2006) suggests “using a family of measures to assess organizational performance,” by using a group of four to six measures that provide managers with an overview “health check” of the areas of interest. A family of measures can be weighted and balanced to give a combined overall score of the company or project. The weight of each measure is not intended to show which measures are the most important, but rather help focus on which measures are most influential on acceptability or the quality process. For instance, if your most important measures are currently meeting requirements then they should not carry the most weight. A less important measure that has not been meeting requirements should be weighted higher in order to focus on improving that area. Thor encourages companies to keep track of their measures over time so that corrective actions can be correlated to changes in performance which may provide some insight for future actions.

Similar to Thor’s family of measures, Kaplan and Norton (1992) propose a balanced scorecard to measure the performance of a project. A balanced scorecard examines multiple metrics within four categories: internal, external, customer and innovation and education. They cite the following benefits of the balanced scorecard:

- Allows companies to decide what to focus on and measure;
- Consolidates management’s focus;
- Allows a holistic view and reduces sub-optimization;

- Focuses attention into four categories of customer concern: time, quality, performance, and cost; and
- Gets senior management involved with quality improvement.

Balanced scorecards have successfully been implemented in companies such as Hewlett Packard, the U.S. Department of Energy, and Philips Electronics for well over a decade.

Another rich source of quality metric information is from American Society for Quality (ASQ). Recently, ASQ published a comprehensive report entitled “The Global State of Quality Research Discoveries 2013” that outlines findings of worldwide practices used by organizations to govern, manage, measure, and support quality in organizations. The report documents the lack of research in the use of tools and techniques in quality management that could provide a comprehensive view of the current state and thus identify future opportunities. The goal is to provide a baseline of these techniques that will support ongoing research into the use quality tools in organizations. The ASQ research (2013) examines four major research questions:

- What is the best governance and management structure for quality to maximize the impact on outcomes?
- What measures of quality should an organization use to drive value?
- How should quality be supported by an organization?
- How can we affect the culture of quality to make it change the way we work?

Findings suggest that:

- There are significant differences in the use and application of quality practices between manufacturing-focused and service-based organizations. This includes governance and management models, the availability of and use of metrics, quality

management frameworks and certifications, and training. In general, manufacturing organizations are more likely to utilize mature quality practices.

- There is a common notion that larger organizations (based on annual revenue) tend to use more mature quality practices than counterparts in smaller organizations.

Although this statement is true for several practices, in general the size of an organization has a much smaller impact on variability in the application of mature quality practices than an organization's specific industry.

- There is no significant indication that the use of quality practices generally differs by region. A few variations do exist but are typically related to size, industry, or other unidentified factors.

This analysis conducted by ASQ on various company sizes, focuses, and locations in reference to quality practices contributed extensively to the development of quality metrics for the capital facilities delivery industry. The findings of this report helped frame the areas to investigate in the literature.

The American Productivity & Quality Center (APQC) is also tackling the collection of quality metrics in a study titled "Enterprise Quality Process Framework" (APQC, 2014). APQC has identified a list of common quality questions that help define the quality process:

- How much money is spent on quality?
- What customer value does that quality expenditure create?
- Is there a 'right' balance between the push and pull of quality process standards?
- Is there a 'right' way to govern and manage organizational quality?
- How is quality done? Where is it done?
- What is quality?

- What quality competency does the organization need based on the current maturity?
- How can quality be benchmarked across organizations and industries? And if it can, how can an organization determine which parts of the quality system have opportunities for improvement?
- How can the expenditure on quality practices be linked to the performance of core business measures?

The study describes a quality process consisting of (1) establishing quality requirements, (2) evaluating performance to requirements, (3) driving improvement, and (4) managing non-conformances. These four processes can be mapped onto the enterprise across the lifecycle consisting of research and design, procurement, production and customer service. The quality process supports common quality measures, which support a quality performance index, allowing for quality benchmarks and quality practices. APQC is still working to complete this project.

In summary, this review of general quality metrics and frameworks leads us to conclude that metrics should assess the level of conformance to defined requirements and help answer any quality questions that a company may have. They should also be few in number so that quality can be assessed, but not be overly complicated. Finally, quality metrics should be used to drive improvement. In reference to the goals of this research project, the metrics that are developed will also need to be used to compare the quality performance across multiple companies. For this reason, these metrics will need to be normalized.

2.2.1. Normalization

Normalization is the process of providing a common basis to compare metrics of different organizations. It can be used to transform data of varying scales and magnitudes into

comparable measures. For example, capital facilities projects have varying costs and size, so quality metrics could be transformed using project cost or labor hours. This will create standardized metrics that can be used for any sized project within an organization or across organizations.

2.3. Quality metrics outside the capital facilities delivery industry

Quality metrics are examined outside the capital facilities delivery industry including healthcare, food and drug, Department of Energy, and the Occupational Safety and Health Administration (OSHA).

2.3.1. Healthcare

Buttell et al (2008) discuss concepts and practices of quality within the healthcare industry. Based on their definition, “Quality consists of the degree to which health services for individuals and populations increase the likelihood of desired health outcomes (quality principles), are consistent with current professional knowledge (professional practitioner skill), and meet the expectations of healthcare users (the marketplace).” Note that the author was careful to include three key factors of quality; quality principles, professional practitioner skill, and marketplace satisfaction. This means that quality in healthcare strives to continuously improve their deliverables, remain up-to-date on the latest technologies and information, and meet the requirements of their customers.

The Institute of Medicine’s Committee on the Quality of Healthcare in America released a publication that outlined exactly how they intend to drive quality improvement within the healthcare industry. First, they defined all parties involved in the effort and recommended that they commit to a statement of purpose for healthcare. This purpose would serve as a new culture

for the healthcare industry to keep quality in mind as they conduct their normal processes. The committee proposed six components that define quality in healthcare (Buttell, p.68).

- *Safe: Avoiding injuries to patients from the care that is intended to help them.*
- *Effective: Providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding underuse and overuse, respectively).*
- *Patient centered: Providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions.*
- *Timely: Reducing waits and sometimes harmful delays for both those who receive and those who give care.*
- *Efficient: Avoiding waste, including waste of equipment, supplies, ideas, and energy.*
- *Equitable: Providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status.*

The analysis provided by this reference of the systematic approach taken by the healthcare industry to improve their quality of care as well as an outline of lessons learned will serve as a great framework for this research.

2.3.2. Pharmaceutical

On July 12, 2013, the Director of the FDA, Howard Sklamberg, announced a change in the way the FDA addresses quality in pharmaceuticals due to the poor quality of products distributed in the U.S. Specifically, the pharmaceutical industry has not had an easy way to evaluate the level of quality across the industry. Recalls and consequent drug shortages have occurred, reflecting poorly on the administration. The FDA and the industry in general have a desire, similar to

RT313, to develop metrics to improve the quality of their deliverables. Their goal is to use quantitative metrics to assess the quality control of their facilities and build a strong quality management system based on the following principles:

- Establish clear standards for review and inspection
- Clear enforcement policies
- Same standards for all drugs; lifecycle approach
- Specialization and team review; integration of review and inspection for a quality assessment
- Clinically relevant standards
- Surveillance using quantitative metrics
- Overall Quality Management System (QMS) and evaluation system

The FDA and the Parenteral Drug Association (PDA) hosted a conference of SMEs in quality metrics and systems in the pharmaceutical industry. This served as a brainstorming session to assist in developing quality metrics and a framework. The following topics were discussed in breakout sessions:

- Executive Level on Metrics
- Product Quality Metrics
- Site and Quality Systems Metrics
- Culture and Comparing Metrics

Over 300 academic and professional individuals attended the conference where at least two SMEs on each category discussed on how they believe the FDA should approach the project. Voting was used during the breakout sessions to gain consensus on important topics and proposed quality metrics and explore options for refinement. Many of the metrics receiving a

high number of votes were very industry specific however, metrics such as confirmed product quality complain rate, recall rate, batch failure rate, and critical investigation rate were consistence with the concerns and discussion of RT313 (and the subject of this master’s thesis) and validate that metrics such as these could be useful in the capital facilities delivery industry. Table 2-2 provides a complete list the metrics receiving the most votes in each category.

Table 2-2. Pharmaceutical Quality Metrics (Mendivil et al, p. 544-545)

Product Quality Metrics	Site and System Quality Metrics
1. Confirmed Product Quality Complaint Rate	1. Confirmed OOS rate (DS&DP)/# tests by site
2. Health authority Quality Notifications (i.e., FAR/BPDR, Recall)	2. CAPA effectiveness
3. Confirmed OOS rate (DS & DP) / # of spec tests by Product	3. CAPA timeliness
4. Process Capability by Product (Cpk, Ppk, etc.)	4. Process capability metric (i.e., CpK, PpK, etc.)
5. Recall Rate by Product	5. Batch failure rate
6. Out of trend by Product for Drug Substance and/or Drug Product	6. Confirmed OOT rates by site (exceeding an action level)
7. Critical Investigation Rate	7. Critical Investigations / Deviations Rate by site
8. Line Clearance Failures	8. Training Effectiveness / On-time Completion
9. Percentage successful effectiveness checks of Drug Substance specific CAPA-correct true root cause first time	9. Percentage of overdue PM for critical equipment
10. On time Drug Substance CAPA closure	10. PIC/S Inspection Scoring. Number of PIC/S member inspections and number of critical & major observations
11. Adverse Event rate/ Safety Signal Detection	11. Environmental monitoring Grade A&B areas (excursions)
12. Confirmed OOS	12. Analytical invalid rate
13. Batch Rejection Rate	13. Recapitalization and PM as a % of the asset value commitment
14. On-time Stability Testing	14. Risk Management & Mitigation Profile Changes
15. Finish product Reject Rate	15. Cycle times (disposition and end to end)
16. Batches Rework/Reprocess Rate	16. Audit/inspection commitment on time completion dates

The conference summary report states that quality metrics should consider three important perspectives: patient, regulator, and manufacturer (PDA, 2014). Some further steps that can be taken to mature the list of metrics identified at the conference include the following:

- Discussing quantitative (generally more objective) metrics vs. qualitative metrics (often with few shared definitions across industry) and developing common definitions;
- Discussing the difference between product specific vs. site and system quality metrics;
- Defining the difference between GMP metrics to indicate compliance with regulations vs. metrics beyond GMP which assess overall quality status of a site or product and the quality culture;
- Identifying and defining leading metrics (generally harder to define but more useful) vs. lagging metrics; and
- Discussing the use of external (commonly understood and reportable to regulators) vs. internal metrics (defined by a company or site and used diagnostically).

The PDA also conducted blind surveys in order to identify potential issues and consequences associated with the recommended metrics. The methodologies used were evaluated and used to identify which approaches will be best given the project's goals.

In December of 2014, the PDA hosted a follow-on conference to provide an update to the industry and conduct further research in the same brainstorming format. Updates from the FDA indicate that the organization is moving towards a rating system that includes a severity scale for the level of quality performance. Once the program is in place, the FDA mentioned the added benefit that companies could use these metrics to assist in contract acquisitions. The organizations also acknowledged a necessity to scale (or normalize) the metrics when comparing them across projects and across companies in order to compare them in similar terms. All of

these characteristics of a QMS that are actively being discussed by the FDA are also characteristics of a QMS that have been investigated and discussed during this study for the capital facilities delivery industry.

Finally, the PDA conference itself focused on the need for a quality culture to be implemented across the industry as well. SMEs from across the pharmaceutical industry presented work on the quality culture and the possibility for collecting metrics on this aspect as well. There was an overwhelming agreement that there is currently a culture of resistance toward reporting quality metrics in fear of a negative result. Being able to assess the actual current state of quality, with minimal lack of reporting, is instrumental in improving quality performance. This will allow issues to be identified and improved so that further quality events do not occur in the future rather than fixing issues as they come along. In the end, improving the culture of quality is thought to have a great impact on the cost of poor quality.

One key difference between the food and drug industry and the capital facilities delivery industry is that the latter does not have a central regulatory organization that sets forth quality requirements and regulations to which all organizations must adhere. To this point, the FDA requires all organizations under their jurisdiction to develop a QMS that is maintained in all levels of the company. Under this regulation, which can be seen on the FDA website (FDA, 2013), organizations are required to compile documentation on their quality procedures and control plans and submit to a quality audit. Corrective actions are required for any operations found to be out of regulation. A written report is to be submitted to management for review. Suppliers, contractors and consultants who supply purchased materials are also required to conform to quality specifications. It is the organization's responsibility to conduct controls and documentations to ensure that any purchased materials are of sufficient quality.

2.3.3. Department of Energy

The Department of Energy (DOE, 2009) began the process of implementing safety improvement standards within their organization in 1996 with the addition of Integrated Safety Management (ISM). By integrating the core functions and principles of ISM, the DOE strives for a *proactive safety culture where:*

- *Facility operations are recognized for their excellence and high-reliability;*
- *Everyone accepts responsibility for their own safety and the safety of others;*
- *Organization systems and processes provide mechanisms to identify systematic weaknesses and assure adequate controls; and*
- *Continuous learning and improvement are expected and consistently achieved.*

The ISM guiding principles used to develop this culture are as follows (DOE, 2009):

1. *Line Management Responsibility for Safety. Line management is directly responsible for the protection of the public, the workers, and the environment.*
2. *Clear Roles and Responsibilities. Clear and unambiguous lines of authority and responsibility for ensuring safety shall be established and maintained at all organizational levels within the Department and its contractors.*
3. *Competence Commensurate with Responsibilities. Personnel shall possess the experience, knowledge, skills, and abilities that are necessary to discharge their responsibilities.*
4. *Balanced Priorities. Resources shall be effectively allocated to address safety, programmatic, and operational considerations. Protecting the public, the workers, and the environment shall be a priority whenever activities are planned and performed.*
5. *Identification of Safety Standards and Requirements. Before work is performed, the associated hazards shall be evaluated and an agreed-upon set of safety standards and*

requirements shall be established which, if properly implemented, will provide adequate assurance that the public, the workers, and the environment are protected from adverse consequences.

6. *Hazard Controls Tailored to Work Being Performed. Administrative and engineering controls to prevent and mitigate hazards shall be tailored to the work being performed and associated hazards.*
7. *Operations Authorization. The conditions and requirements to be satisfied for operations to be initiated and conducted shall be clearly established and agreed upon.*

ISM core functions that accompany these principles that help develop the culture are as follows (DOE, 2009):

1. *Define the Scope of Work. Missions are translated into work; expectations are set; tasks are identified and prioritized; and resources are allocated.*
2. *Analyze the Hazards. Hazards associated with the work are identified, analyzed, and categorized.*
3. *Develop and Implement Hazard Controls. Applicable standards and requirements are identified and agreed-upon; controls to prevent or mitigate hazards are identified; the safety envelope is established; and controls are implemented.*
4. *Perform Work within Controls. Readiness to do the work is confirmed and work is carried out safely.*
5. *Provide Feedback and Continuous Improvement. Feedback information on the adequacy of controls is gathered; opportunities for improving how work is defined and planned are identified and implemented; line and independent oversight is conducted; and, if necessary, regulatory enforcement actions occur.*

The DOE's systematic approach to developing their safety management program is yet another source in which the team can pull from to customize their approach to developing quality metrics in a way that will maximize success and satisfy stakeholders.

2.3.4. Safety Metrics

In 1931, H. W. Heinrich published a book called "Industrial Accident Prevention" that investigated the relationship between different safety events. He was the first to notice a trend between near misses and injuries, claiming that for 1 major injury, there are 29 minor injuries and 300 near misses involving the same person. To illustrate this finding, Heinrich created a pyramid diagram, shown in Figure 2-1 that featured near misses, minor injuries, and major injuries as layers, building in severity.

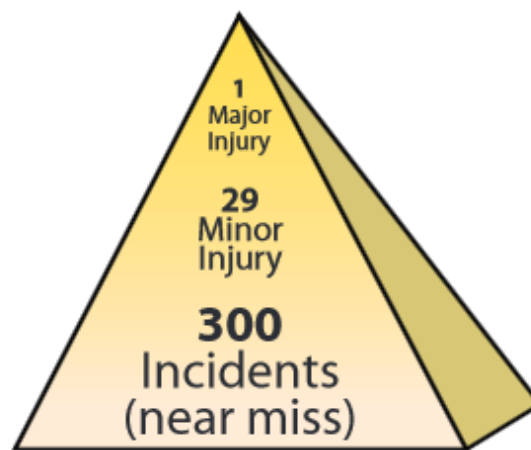


Figure 2-1. Heinrich's Safety Pyramid (Heinrich, 1959)

Years later, in 1969, Frank E. Bird Jr. wanted to determine a ratio, similar to this, that represented the average population of workers. Bird conducted a study that included 1,753,498 accidents reported by 297 companies. This study reported similar findings, namely that for every reported major injury, there were 9.8 minor injuries reported. ConocoPhillips Marine (Masimore,

p.11) built on this study by expanding the types of safety events and including a base level of “unsafe” or “at-risk” behaviors. This pyramid is widely used today, with the events that make it being collected by OSHA.



Figure 2-2. Safety Pyramid from ConocoPhillips Marine (Masimore, p.11)

The adoption of this pyramid framework by a central organization such as OSHA allows for the creation of the safety culture within companies across various industries which led to the uniform improvement of safety within the workplace. This research to develop a quality metric for the capital facilities delivery industry modeled this safety approach.

2.4. Quality Metrics for the Capital Facilities Delivery Industry

2.4.1. Quality Metrics in the US

Currently, the “Benchmarking and Metrics Program” from CII is the main quality metric initiative used by the capital facilities delivery industry. “It aims to provide performance norms to the industry, quantify the use and value of best practices, and help focus CII research and implementation efforts (Costa, p. 458).” The most popular metrics used under this program include the following:

- Budgeted and actual project costs;
- Planned and actual project schedule;
- Facility capacity;
- Work hours and accident data; and
- Project impact factors.

2.4.2. Quality Metrics in other Countries

Costa (2004) summarizes the most widely accepted performance measurement techniques of four different countries; Brazil, Chile, the UK, and the U.S. Table 2-3 depicts a summary of these strategies examining common metrics and cited issues.

Cha (2011) conducted a thorough analysis of the capital facilities delivery industry both domestically and globally, in order to develop a more effective performance measurement system for construction projects in South Korea. After analyzing relevant metrics used in the industry globally, 18 top indicators were identified and relative weights were assigned. If all 18 indicators were measured and used in a scorecard format, the resulting total score would reflect the overall performance of the project executors. The authors also recommended that the metrics are tracked from project to project and used to predict the success rate of future projects. If organizations across the industry chose to adopt this performance measurement system, it could be used for contractor selection. The indicators resulting from this study are listed in Table 2-4.

The study also found that, based on the current analysis of the industry, the majority of the metrics or indicators are not typically used during project execution; rather they are measured and analyzed later. The typical measures only assess the efficiency of cost and schedule rather than the actual quality of the product being built. A set of metrics that provide an indication of

the quality of the product being delivered would allow for an entirely new comparison to be made between competing companies.

Table 2-3. Performance Metric Programs from Various Countries.

Program and Country of Practice	Measures	Issues
Key Performance Indicators (KPI) from the UK	<ul style="list-style-type: none"> • Client Satisfaction-product • Client satisfaction-service • Construction cost • Construction time • Defects • Predictability-cost • Predictability-time • Profitability • Productivity • Safety 	<ul style="list-style-type: none"> • Project specific • Offer little indication on performance of organization • Many indicators can be linked together to form other viewpoints • It is up to the user to input to correct measurements • No indicators for suppliers • Other concerns with availability of data and the indicators validity • No action/reaction plan.
National Benchmarking for Chile	<ul style="list-style-type: none"> • Deviation of cost by project • Deviation of due date • Change in contract • Accident rate/Risk rate • Efficiency of direct labor • Productivity performance • Rate of subcontracting • Client cost complaints • Urgent orders • Planning effectiveness 	<ul style="list-style-type: none"> • Program assumes that the company already has a QMS in place to assist in measuring indicators. Because of this, indicators were not always easy to measure. • Lack of commitment to benchmark at company level • Poor standardization of measures
Performance Measure System for Brazil System of Quality and Productivity Indicators for the Construction Industry (SISIND)	<ul style="list-style-type: none"> • Cost deviation • Time deviation • Non-conformity index for critical processes • PPC • Supplier performance • Degree of user satisfaction • Sales time • Ratio between the amount of accidents and total man-hour input • Construction site best practice index • Degree of internal client (workers) satisfaction 	<ul style="list-style-type: none"> • Customers tend to collect some indicators that are not related to critical processes only because those are easy to collect. • Companies found the indicators with both quality and quantity indicators were hard to implement. • Lack of human resources • Companies that failed to establish priorities for their indicators failed to seeing benefit from program because no improvement actions were defined.

Table 2-4. Quality Metrics summarized from Cha (2011)

<ul style="list-style-type: none">• Cost Efficiency• Cost Effectiveness• Construction Cost Predictability• Schedule Efficiency• Construction Schedule Predictability• Time Savings• Defect Frequency• Rework Frequency	<ul style="list-style-type: none">• Accident Rate• Safety Cost Ratio• Safety Education• Site Dangerousness• Construction Waste Rate• Recycling Rate• Management Productivity• Labor Productivity
---	---

2.5. Conclusion

All sources/industries agree that the first step to improving quality is a culture change usually beginning with top leadership. They have also recognized the benefit of employing metrics can illustrate change over time. There is an emphasis on clearly defining the roles and responsibilities of all members involved in the project lifecycles so that nothing is left out or overlapped. Lastly, one major factor in the success of these industries implementing quality metrics into their organizations is the existence of a higher regulator system enforcing or encouraging them to do so. Many sources have anticipated a loss in motivation otherwise.

Next, Chapter 3 describes the approach and methodology that was used to develop the deliverables of this research study as well as the tools used for data collection and data analysis.

3. Methodology

3.1. Introduction

This chapter outlines the methodology used to achieve the thesis research objectives in support of the ongoing CII research project being conducted by RT313. The primary objective and motivation for this research is to create a set of quality metrics for the capital facilities delivery industry that will allow both owners and contractors to assess the quality performance of their projects, identify areas of improvement, and demonstrate quality performance relative to the industry, for both internal and external customers, and use the performance indicated by the metrics to differentiate themselves. The approach taken to carry out this research included the following:

1. Replicate as much as possible the characteristics present in today's accepted safety metrics;
 - Safety metrics and the safety pyramid are the one uniformly accepted practice and measurement of performance that the capital facilities industry has today. Furthermore, safety has obvious comparisons to quality, for example, some safety events are more severe than others and preventative actions should be taken in attempt to prevent safety events from happening. Because of this, replicating the widely-accepted and successful safety programs to create a quality program not only makes sense from a qualitative standpoint, but also will increase the likelihood of acceptance and understanding of a groundbreaking new quality program such as the one the RT313 team is attempting to create. For this reason, replicating safety was chosen as RT313's first aim in creating a new quality program.

2. Establish a common understanding (common language relative to quality similar to the common language utilized in measuring safety performance);
 - A common language and understanding of a uniform quality program is necessary to ensure truly uniform adoption. If companies were to implement quality metrics in different ways, it would lead to inaccurate comparisons.
3. Define what a quality failure is;
 - Currently, there is no uniform definition of a quality failure within the capital facilities delivery industry. Some companies see a quality failure as any deviation from the design requirement while other companies would see a quality failure as an event that would cause the project as a whole to fail. These definitions are vastly different. In creating quality program that spans the entire industry, it would be necessary to gain consensus on what a quality failure is.
4. Benchmark current performance and identify trends both internally and across the industry;
 - Being able to benchmark performance measures obtained from quality metrics will allow companies to understand how they compare to the industry as a whole. Also, when used internally, companies can compare their projects to each other. This will allow companies to identify areas of improvement within their quality performance.

The target population for this study consisted of companies that are members of CII that had a pre-existing quality program and associated quality management system in place that would allow the research team to obtain quality data on current and past projects. The initial pilot study was limited to those companies that employ the Subject Matter Experts (SMEs) on RT313.

Subsequent studies were broadened to include the CII Quality Management Community of Practice (QMCOP) and the CII membership, at large. Both sampling populations contain both owner and contractor companies.

Project quality events were identified and defined by the research team based on the scope of quality for the capital facilities delivery industry. Quality event data was collected from current (open) and past (closed) projects, using a data collection tool designed by the research team, so that the overall quality could be assessed on a project-to-project basis, rather than other possible reporting methods such as annual, regional, or companywide. Once the program was completely integrated into a company's quality management system (QMS), the goal would be to use a companywide metric for quality. This chapter will illustrate the methods used to develop the framework for this quality program, including the definition of quality, the development of the quality pyramid, and the data collection tool and data analysis.

3.2. Quality Framework

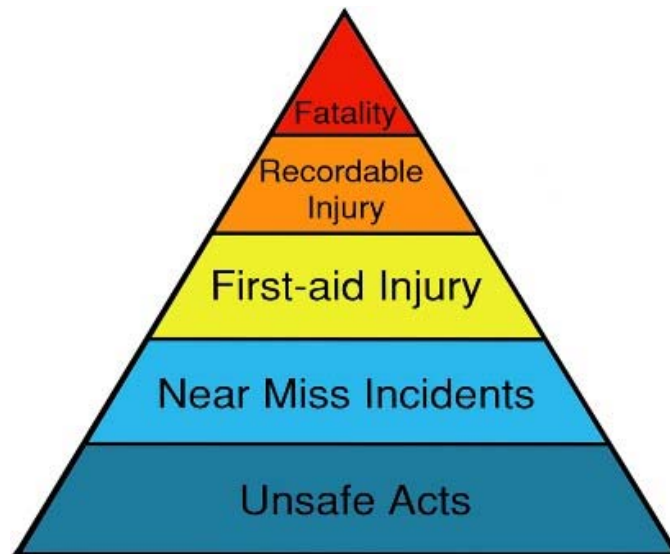
3.2.1. Quality Definition

In order to develop quality metrics for the capital facilities delivery industry, a consensus needed to be reached on the definition for quality as it pertains to projects in the industry. A literature review was conducted to gain knowledge on how quality is traditionally defined by other industries and quality experts. Recall in Table 2-1 the set of widely accepted definitions of quality. These definitions are not limited to the capital facilities delivery industry; rather they are definitions of quality that are accepted across all industries. This table was used during the October 2013 RT313 meeting as a reference to assist in defining project quality. SMEs also brought definitions of quality from their own companies. All definitions were considered and discussed in order to determine a definition that best represented the needs of the capital facilities

delivery industry. In this situation, the industry needs a definition for the quality performance of the actual deliverable. Having a definition that addressed this form of quality would help build the foundation in which the quality metrics were built upon.

3.2.2. Quality Pyramid

Due to the widely accepted nature of the safety pyramid, the team was tasked with creating a quality pyramid that would accomplish the same goal. Shown below in Figure 3-1 is an illustration of the safety pyramid from OSHA. The safety pyramid begins with unsafe actions at the base that result in unplanned safety events found in the subsequent levels. Each level up the pyramid represents a more severe safety event; eventually leading to a fatality. OSHA was able to standardize safety metrics across multiple industries using this pyramid framework. These metrics are also found to be consistent with the findings of the literature review; they are few in number, easy to understand, easy to collect and they encompass the scope of interest.



Source: <http://www.enerpipeinc.com/HowWeDoIt/Pages/safety.aspx>
Figure 3-1. Example of the OSHA Safety Pyramid

Following this methodology, the quality pyramid would theoretically have quality *activities* (or leading indicators of quality) at the bottom of the pyramid, followed by quality *events* (or lagging

indicators of quality) that would increase in severity with each layer, moving upward. The team participated in many brainstorming sessions in order to determine a set of quality activities and events. Many SMEs provided input from their experience with quality conformance issues as well as definitions that their companies use to specify different quality events. The team expressed the need for the pyramid to be simple enough, to ensure maximum acceptance, but broad enough to ensure that all severity levels would be represented, no matter the size of the project or company. This general yet broad aspiration proved to be a very tedious process. As a result, the team spent over six months participating in iterations of refining the pyramid until everyone could agree that it encompassed all possible quality events. During this process, event categories were added when further distinctions could be made, event categories were removed when they were too similar or redundant to another, and some event categories were combined when two or more events were sufficiently unique but neither was more severe than the other. One of the goals of this process was to define the events in such a way that actual quality events could not be assigned in different event classifications. This would ensure each event was unique, more severe than the previous, and necessary to span the scope of quality performance.

Once final definitions were agreed upon, additional documents were developed; providing examples that further illustrate each quality activity and event to ensure that data would be collected correctly. These illustrations were drafted from an owner and a contractor perspective and were presented to the research team for discussion. After several iterations of refinement, a set of examples for each quality activity and event were created to further assist companies in the implementation and understanding of the quality program.

3.2.3. Data Collection Tool

Similar to the events in the safety pyramid, the quality events in the quality pyramid would serve as the metrics or indicators for data collection purposes. Originally, an Excel file was created for

the companies to report their data. The first spreadsheet of this file contained instructions on how to record quality event data and a picture of the quality pyramid with event definitions attached. On the second spreadsheet, the user had space to enter company data, such as, industry group (heavy, light, building, infrastructure) and company type (owner, contractor) followed by project data, consisting of project cost, location, start and end date, and project labor hours. Finally, there was space for event data, including event type, a short event description, the cost to correct the quality event, the labor hours it took to correct the event, and the date and lifecycle phase (planning, procurement, design, construction, turnover) in which the event occurred and was discovered.

This data entry spreadsheet would automatically sum each event reported in a project summary table. While creating the tool for pilot data collection, feedback from SMEs showed a lack of wanting or an inability to report certain fields. For example, many SMEs agreed that they were not comfortable with reporting the cost and labor hours needed to address the quality event. In many cases, companies were not even tracking this kind of data and therefore would not be able to report it. The team agreed that these were interesting characteristics to collect and could be used for future improvement purposes but it was ultimately decided that it would not be possible to collect this data for past (closed) projects. It should be noted that if a company were to implement a data collection tool such as the one being developed for this project, they could collect these characteristics for new projects. This led to a stage where project that was primarily focused on refining the data collection tool to include data that all or most companies could report as well as, making it more user-friendly.

The process of developing this data collection tool consisted of several iterations of revisions through discussions with research team members and SMEs in order to ensure the

maximum amount of reporting compliance by companies using the tool. If there were too many fields that companies were unable to report, respondents tended to get discouraged and not report data at all. Therefore, during each iteration of refining the data collection tool, the amount of data input fields was reduced. Input fields were deleted only if they were deemed unnecessary to the scope of the project or if it became apparent that companies would not be able to report the data.

Once all of the data fields were agreed upon, the final data collection tool was developed using survey software called Qualtrics. This survey format allowed data reporting to be simplified greatly. Only fields that were pertinent to the data analysis and metrics reporting were kept in the survey. Initially, two pilot data studies were conducted; first with the SMEs on the research team and second with the members of the Quality Management Community of Practice (QMCOP) group of CII. All of the data that was collected using the Excel tools was entered into the Qualtrics database. The results and feedback from these studies helped refine the tool to the final version. Once this was done, the data collection tool was launched to the entire CII population. The final version of the Qualtrics survey had to be approved through CII before the study could be conducted.

3.3. Data Analysis

Data was collected for this study in three different phases. First, a pilot study using the Excel tool was sent out to the 24 CII members who served as SMEs on RT313. The SMEs were from different organizations and were asked to provide project data for at least five projects.

Seventeen members submitted 51 projects by the deadline. In order to maintain confidentiality of the organization's and the projects information, all the data were stored by the academics so that no one else would have access to it.

Next, the Qualtrics tool was used to collect data from the RT313 members again, along with members of the CII QMCOP. A two month time limit was given and 22 more projects were added as a result of this study, bringing the total number of projects to 73. Ten of these projects were contributed from QMCOP companies. The final data collection study was conducted to include the entire CII membership with a time limit of another two months. Emails were sent out to the CII data liaison's for each company and included a value case statement to explain why they should fill out the study, a PDF attachment of what the survey would look like, and a PDF attachment of instructions on how to collect and report the data. A total of 29 additional projects were reported by the due date, making a final data set 102 projects.

3.3.1. Normalization

In analyzing the data received from the three data collection studies, it became very apparent that projects had a great deal of variation among them. The magnitudes of total project cost, labor hours, project duration, and reported events all varied greatly. Because of this, simply using the raw data, like number of events, would lead to inaccurate comparisons across projects. In these instances of comparing project-to-project or company-to-company, the data was normalized based on characteristics that all projects had in common; cost, time, and work hours. This would allow projects and companies to be compared on a “per dollar,” “per work hour,” or “per unit time” level. By normalizing the data, comparisons on project data could be made despite the varying sizes of projects or companies.

3.3.2. Hypotheses

Five research hypotheses statements were developed in order to test the validity of the quality pyramid and data collection tool from the project level. These statements were first drafted by the academic team then presented to the SMEs. SMEs provided feedback on which aspects of the pyramid needed to be validated in order to gain buy-in from the industry and how they should be

measured. Similar to the process of developing the quality events, these hypotheses statements were refined by the research team, over a period of time, until a consensus was reached.

The first two hypotheses aim to test the relationship between planned activities and unplanned events. Since planned activities are conducted to correct quality deficiencies before they become quality events, the research team hypothesized that the more planned activities that are conducted, the less quality events that should occur. To test this, Pearson Correlation analyses were used. A Pearson correlation is a measure of the linear correlation between two variables. This type of analysis would allow the research team to better understand how planned activities can affect unplanned events by looking at the correlation of these two metrics across all of the projects in the data set. The resultant correlation value, or correlation coefficient, is a value ranging from negative one to positive one, where a positive one represents a complete positive correlation, a zero represents no correlation, and a negative one represents a total negative correlation. In this case, a negative correlation would be desired since planned activities should increase as unplanned events would decrease.

The third hypothesis would test the relationship between self-reported perceptions of quality, reported by quality experts about their company projects, and normalized unplanned events. The idea is that if SMEs were able to effectively assess their project quality using their QMS, the rating would increase (indicating high quality) as the number of unplanned events decreases. Respondents were asked to report the level at which they agree with the following question, based on a seven point Likert scale; this project was a high quality project. Then, a Pearson Correlation analysis was again used to understand the relationship between the variations of the two variables. For this hypothesis, a negative correlation would show that SMEs are able to correctly assess the quality performance of their company projects, while a no

correlation or a positive correlation would show that SMEs were unable to correctly assess project performance. This type of outcome would potentially provide justification for the need for an industry-wide QMS.

Hypothesis 4 also aimed to test the relationship between the self-reported perceptions of quality and unplanned events. However, this test separates the data set into two samples (high and low self-reported quality) and uses a t-test to infer a difference in the mean of each unplanned event. Therefore, projects with low reported quality should have a higher number of events than projects with high quality. The hypothesis statements and test statistic for this test can be seen below with μ_1 and μ_2 being the average number of unplanned events for low self-reported projects and high self-reported projects, respectively.

$$H_0: \mu_1 = \mu_2 \quad \text{Equation 3-1.}$$

$$H_1: \mu_1 > \mu_2$$

$$t_0 = \frac{\mu_1 - \mu_2}{s/\sqrt{n}} \quad \text{Equation 3-2.}$$

Using this type of mean hypothesis test, the p-value would be the sum of the probability distribution above t_0 . Therefore, by employing a confidence level of 95%, we would reject the null hypothesis when the p-value is equal to 0.05 or less. This would allow us to conclude that low self-reported quality projects actually perform worse than high self-reported projects.

Finally, the fifth hypothesis would test the idea of the pyramid structure itself, within each project rather than across projects. The research team hypothesized that for any given project, the proportion of events to total events should be greatest at the base of the pyramid, and then each subsequent event should progressively get smaller. This hypothesis also used statistical

mean hypothesis testing, similar to Hypothesis 4, to infer a difference in the means of the quality events in the quality pyramid. In this case, μ_1 would be the average number of quality events for the first level on the pyramid and μ_2 would be the average number of quality events for the level directly above it, for example. This methodology was repeated until all levels of the pyramid were tested.

3.4. Methodology Conclusion

Descriptive statistics and hypothesis testing were both used to better understand the data that was reported. Input fields found in the survey were summarized in various different ways in order to begin to understand the industries' quality performance using a uniform quality metric. In the end, there were a number of analyses that were not able to be studied due to the industries' inability to report the data that this time. The methodology reported in this paper shows the types of analyses that companies could do when implementing the quality pyramid into their existing QMS.

Next, Chapter 4 presents the results from the pilot data collection efforts including the developed set of quality metrics and frameworks and the outcomes of the data analysis.

4. Results

4.1. Introduction

This chapter outlines the outcomes of this research project. The final version of the definition of quality, quality pyramid, and data collection tool will be discussed. Results from the data collection study will also be presented.

4.2. Quality Framework

4.2.1. Quality Definition

Based on the literature review of widely accepted definitions of quality and discussions on what quality means for the capital facilities delivery industry with SMEs from the RT313 members, experts at their companies and members from the CII Quality Management Community of Practice, the research team came up with the following definition of quality; *Quality is conformance to defined requirements*. The RT313 team noted that safety, cost, schedule and implied items are not included in the research scope. Elements of a project include: safety, cost, schedule and quality, but our study is limited to defining quality metrics. This definition was used as a basis in determining further aspects of quality, such as the development of the quality pyramid and quality events, required as part of this project.

4.2.2. Quality Pyramid

Shown below in Figure 4-1, is the final version of the quality pyramid (Needy et al, Forthcoming). At the base of the pyramid are two leading indicators or planned activities – Prevention and Appraisal. Next, there are four lagging indicators or unplanned events listed in increasing levels of severity including – Findings, Variations, Defects and Failures.

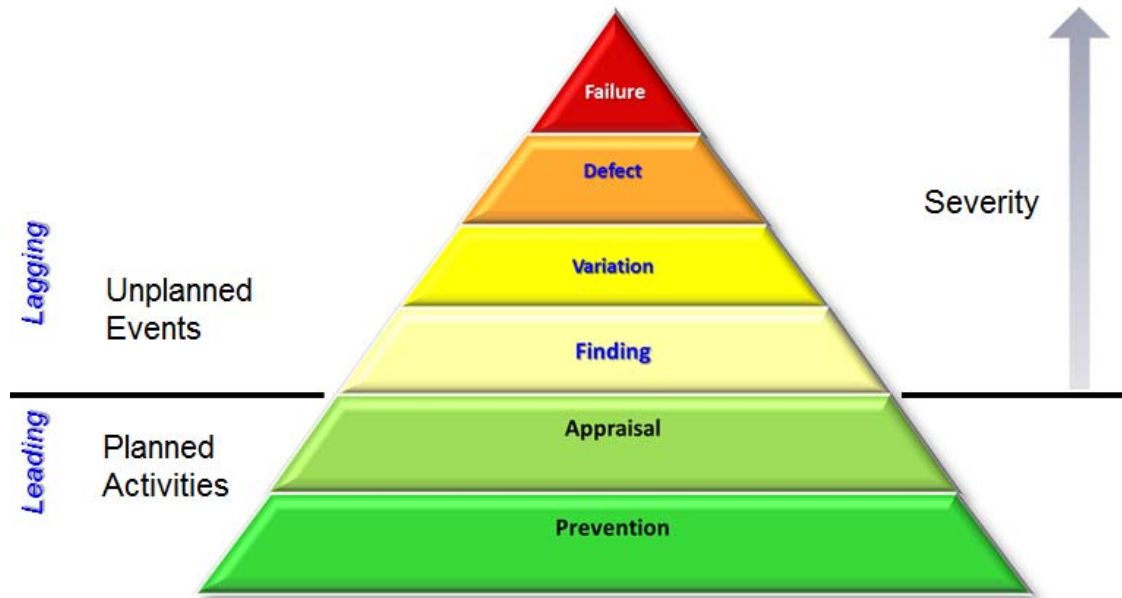


Figure 4-1. *Quality Pyramid*

Detailed definitions have been developed for each severity level, along with descriptions and examples. Definitions for these quality events can be seen in Table 4-1. An illustration document that provides further examples of each event can be found in the Appendix A (Damle, Forthcoming).

Table 4-1: *Quality Event Definitions*

Event	Definition
Prevention	Activities for preventing non-conformance in products or services through the use of established procedures, instructions and/or drawings and other applicable documents.
Appraisal	Planned and documented activities performed to determine the adequacy of, effectiveness of, and compliance with, established procedures, instructions and/or drawings and other applicable documents.
Finding	An outcome of an "Appraisal" that provides documented determination that either the compliance with or implementation of established procedures, resources, instructions or other applicable documents is inadequate. A Finding requires a documented response as to corrective and/or preventive actions with follow-up to assure implementation of stated actions.
Variation	A non-conforming item (system, structure, component or service) that has been determined through investigation, examinations or evaluation of objective evidence that can be accepted as is and may change the requirement.
Defect	A non-conforming item (system, structure, component or service) that has not fulfilled an intended requirement or reasonable expectation for use and requires correction.
Failure	The inability of a system, structure, component or service to perform its required function on demand. Not detected or detectable within the quality program before utilization for intended purpose.

These unplanned events were used as the quality metrics for reported projects by converting them into a Quality Performance Rate (QPR). This was done by dividing the number of events (N), either separately or in total, by the total labor hours expended for the project then multiplying by 200,000 labor hours, as shown in Equation 4-1. Multiplying by 200,000 labor hours is done to be consistent with safety practices (it represents approximately 100 person-years) and also to value to a more readable metric (eliminating a very small decimal value). Recall from Chapter 3 that labor hours was one of the normalizers that was investigated during the study. The RT313 team concluded that labor hours would be the most appropriate normalizers for quality metrics, primarily because human error is typically the cause of quality events occurring during the project lifecycle.

$$\frac{N(\text{Number of quality events}) * 200,000 \text{ labor hours}}{\text{Total Labor Hours Expended}} \quad \text{Equation 4-1.}$$

4.3. Data Collection Tool

In an effort to simplify the data collection process for the user, the tool was redesigned from a set of Excel spreadsheets to a single Qualtrics survey. Qualtrics is an online survey tool that would allow users to input their data and submit it to a database specifically for the RT313. For the final data collection study, a link to the survey was sent out to all of the CII Data Liaisons, as well as, RT313 and QMCOP membership.

The survey link would first lead the user to a consent form that would explain both the study and the confidentiality agreement, shown in Appendix B. If the user agreed to complete the study, the survey would continue to an instruction section, shown in Appendix C, where users were provided with a motivation for the study, quality events and activities definitions, and a check list containing all of the information that they would need to complete the survey. The user

could continue to the actual survey, shown in Appendix D, once all of the required information was available. By using the Qualtrics software, the academics were able to run a report at any time and generate a single spreadsheet containing all of the reported data to review the data completion status and to facilitate detailed data analysis.

4.4. Data Analysis

During the data collection process, respondents were asked to provide information regarding their respective company, including, the name, company type (owner/contractor), and personal contact information. More specific questions were asked about the capital project being reported such as, industry type (heavy, light, building, infrastructure), project status (open/closed), total project cost, location, start and end dates, total project labor hours, and the frequencies of quality events (findings, variations, defects and failures). The analysis of these data could effectively be separated into three sections: demographic data, metrics summary, and hypothesized data.

Various demographic data was calculated and summarized into tables or graphs to get a better understanding of what data companies were actually able to submit. Tables and graphs were also constructed to provide an overall picture of the quality metrics that were developed. Finally, the hypotheses that were developed by the research team were analyzed using statistical testing. These three types of data will be discussed in this chapter.

4.4.1. Demographic Data

Demographic data was collected from the submitted projects in order to understand the types of responses that were being received and to help identify possible areas of further analysis. Below is the most general form of demographic data that was collected. In the initial phases of the data collection process, RT313 companies were asked to submit five to ten projects each for an

estimated sample size of 100 to 200 projects to be used for analysis. Table 4-2 shows the realized sample size or the number of projects that were submitted and from which company type.

Table 4-2. Number of Submitted Projects

Company Type	Number of Projects	Percent of Total
Owner	46	45%
Contractor	56	55%
Total	102	100%

This desired number of projects was not reached until the third and final round of data collection which went out to the general CII population of companies. Table 4-3 shows the number of submitted projects from each section of CII that was requested to submit data; RT313, the QMCOP, and the general CII population of companies.

Table 4-3. Number of Submitted Projects

Respondents	Number of Projects	Percent of Total
General CII	21	21%
QMCOP	14	14%
RT313	67	66%
Total	102	100%

The next two tables show some further industry related demographics with the number of projects in each project size and industry type. Table 4-4 indicates that about half of all submitted projects were considered large, ranging from \$10million to \$1billion, while Table 4-5 indicates that roughly 73% of submitted projects belong in heavy industry.

Table 4-4. Number of Submitted Projects per Project Size

Project Size	Owner	Contractor	Total
Mega (greater than \$1B)	3	13	16
Large (between \$10M and \$1B)	27	30	57
Small (less than \$10M)	16	13	29
Total	46	56	102

Table 4-5. Number of Submitted Projects per Industry Type

Industry Type	Owner	Contractor	Total Projects
Heavy Industry	30	44	74
Light Industry	10	6	16
Building	4	0	4
Infrastructure	0	0	0
Not Reported	2	6	8
Total	46	56	102

Table 4-6 shows some very interesting findings. This table shows the number of projects that utilized various prevention activities. Firstly, there is a great difference between the numbers of prevention activities utilized by owners versus contractors. Based on the data, contractors use more prevention activities than owners. When averaging the total number of prevention activities per project, you find that on average, contractors participated in 1.5 more prevention activities than owners. In particular, contractors seem to use corporate quality metrics, quality infrastructure, and quality training activities approximately twice as often as owners. Discussions from SMEs suggest that this could be due to many owner companies placing the responsibility for prevention activities onto the contractor organizations.

The graph below illustrates this prevention activity data based on the percentage of owner and contractor projects that utilized them. The number of each bar corresponds to the number of each activity in Table 4-6.

Table 4-6. Number of Projects per Prevention Activity

No.	Prevention Activities	Owner (N=46)	Contractor (N=56)	Total (N=102)
1	Scope and criteria management systems are implemented	44	49	93
2	A project execution plan includes a project quality plan for the project	37	52	89
3	Quality roles and responsibilities are documented	32	54	86
4	Quality procedures are implemented	32	53	85
5	Stakeholder alignment has been achieved	39	42	81
6	Contract partners have implemented quality programs	42	33	75
7	Corporate quality metrics and/or procedures are implemented at the project level	19	54	73
8	A quality infrastructure is implemented within the project	20	49	69
9	Quality training activities are in place and have been executed	20	45	65
	Total	285	431	716

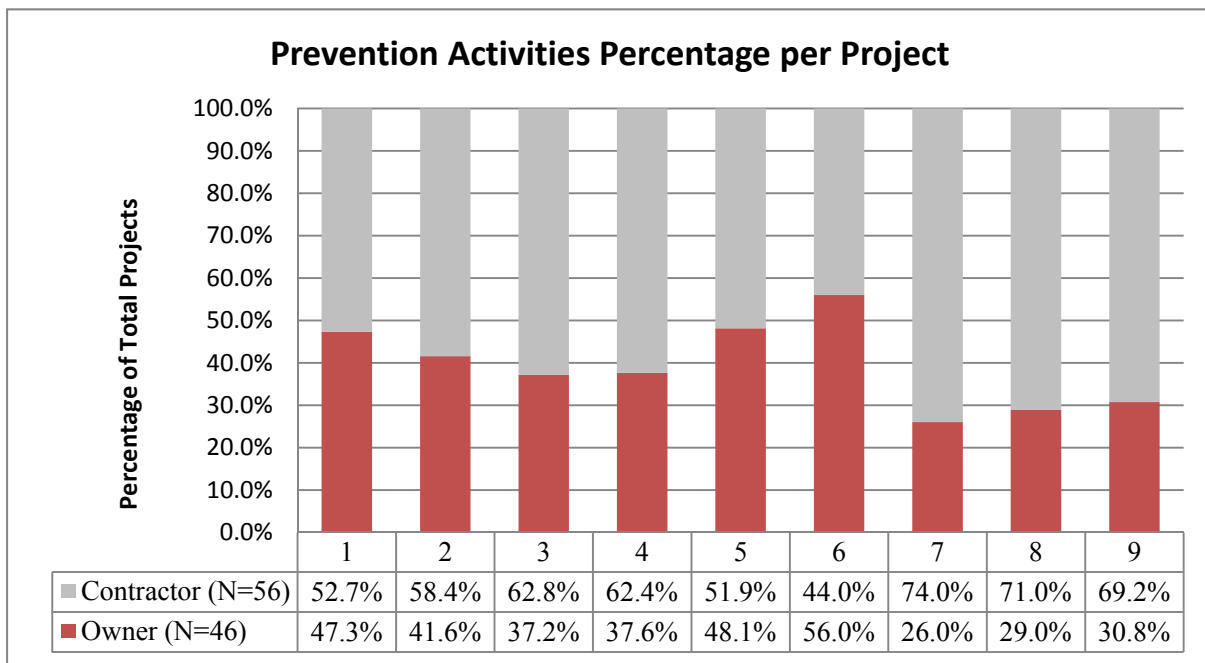


Figure 4-2. Percentage of Prevention Activities used per Project

The final set of demographic type data that was presented is shown in Table 4-7. This table shows the responses to the self-reported perception of quality question found in the survey. In

order to “self-report” their perception of quality, respondents had to rate their level of agreement with the statement, “This is a high quality project” according to a seven-point Likert scale ranging from strongly agree to strongly disagree.

This question was not only used for statistical data analysis, but also to understand how industry members currently feel about the quality within their projects. A graph illustrating this data is also provided in Figure 4-3. This graph helps to show that many of the respondents agreed that their project was a high quality project. In fact, only 10 projects out of the total 102 projects had results showing a disagreement with the statement on some level.

Table 4-7. Self-Reported Perception of Quality

Responses	Owner	Contractor	Total
Strongly Agree	6	14	20
Agree	27	23	50
Somewhat Agree	4	12	16
Neutral	5	1	6
Somewhat Disagree	2	3	5
Disagree	2	0	2
Strongly Disagree	0	3	3
Total	46	56	102

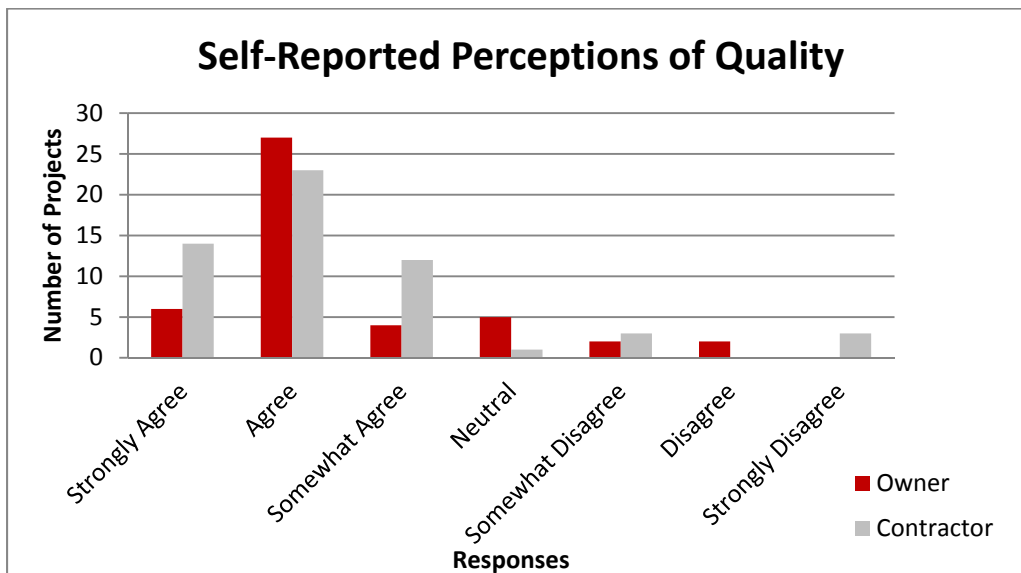


Figure 4-3. Self-Reported Perceptions of Quality Graph

4.4.2. Metrics Data

The next set of data to be reported is the summary metrics data. The number of quality activities and events were reported by owner, contractor, and in total. These numbers were also normalized per project and per labor hour to develop the QPR in order to effectively compare projects of different scales and magnitudes. Some statistical hypothesis testing was also conducted to determine if there was a significant difference in the mean values of some of the reported subgroups. Those tests that were not explicitly defined as hypotheses of the research will be discussed in this section. Specifically, some hypothesis testing was done on demographic information to determine if there was a significant difference between two subgroups.

Table 4-8 shows the average number of appraisal activities and quality events per project, separated by owner and contractor. Prevention activities have been left out of these tests because the survey only asked if a category of prevention was used; the number of times a given prevention activity was used was not reported. This type of data summary helps to show that on a per project basis, contractors conduct more planned activities than owners. It also shows that contractors generally have less quality events per project. While this type of study helps the researchers to understand the type of data that was received, it does not allow for very meaningful comparisons due to the inherent variation across projects.

Table 4-8. Number of Activities/Events by Type and per Project

Activities	Owner (N=46)		Contractor (N=56)		All (N=102)	
	Total	Per project	Total	Per project	Total	Per Project
Appraisal	17,403	378	111,701	1,995	129,104	1,266
Events						
Findings	4,496	98	2,795	50	7,291	71
Variations	433	9	3,269	58	3,702	36
Defects	24,576	534	3,147	56	27,723	272
Failure	103	2	190	3	293	3
Total Events	29,608	644	9,401	168	39,009	382

To better understand the variation between these projects, some descriptive statistics including minimum and maximum values, means, medians, and standard deviations were calculated for each quality event and appraisal activity, shown in Table 4-9 and Table 4-10. For owners, there is a high variability in appraisals, findings, and defects while contractors had a high variability in appraisals. The analysis of activities and events at the project level led to the realization that projects tend to vary in scaling attributes such as project cost, schedule, and labor hours. Because of this, the frequency of the activities and events were then normalized by (divided by) these scaling attributes in order to be able to compare projects in more similar terms.

Table 4-9. Owner Descriptive Statistics for Activities/Events per Project

	Appraisals	Findings	Variations	Defects	Failures
Minimum	0	0	0	0	0
Maximum	4,695	3,478	134	18,575	57
Mean	378.3	97.7	9.4	546.1	2.2
Median	11	1.5	0.5	6	0
Std. Dev.	854.5	510.2	25.4	2,749.8	8.7

Table 4-10. Contractor Descriptive Statistics for Activities/Events per Project

	Appraisals	Findings	Variations	Defects	Failures
Minimum	0	0	0	0	0
Maximum	45,089	322	1,506	646	130
Mean	1,994.7	49.9	58.4	56.2	3.4
Median	36.5	21	0	4	0
Std. Dev.	7,000.5	69.3	219.8	117.0	18.6

To further understand the quality data that was submitted, the appraisals and quality events were summarized using the QPR, as shown in Table 4-11. This means that the total number of each appraisal activity and quality event was multiplied by 200,000 then divided by the actual labor hours expended for the project. This chart shows similar results as the previous one in that it illustrates that contractors perform more appraisals than owners while also having fewer quality events. By using the QPR formula to obtain a unit of the number of events for a project per labor

hour, the quality performance of the set of projects were able to be compared in similar terms for the first time.

It is interesting to point out that while owners conduct roughly one third of the appraisals that contractors do, owners reported an average of 29.5 normalized findings per project versus 7.8 for the contractors. Since these “findings” are defined as the outcomes of an appraisal, it can be concluded that owners tend to have more effective appraisal practices than contractors.

Table 4-11. Number of Normalized Activities/Events by Type

Activities	Owner (N = 46)		Contractor (N = 56)	
	Total	QPR Value	Total	QPR Value
Appraisal	17,403	114.0	111,701	311.3
Events				
Findings	4,496	29.5	2,795	7.8
Variations	433	2.8	3,269	9.1
Defects	24,576	161.0	3,147	8.8
Failure	103	0.7	190	0.5
Total Events	29,608	194.0	9,401	26.2

A further illustration of how companies can be compared can be seen in Figure 4-4. The QPR for each project that submitted labor hours (60 total) was averages for each company and the companies were randomized to create a bar chart so that the overall company metric ratings could be compared. Charts similar to this one help companies see how their overall project quality compares to other companies in the industry.

To better understand the distribution of these normalized events, a scatter plot was created in Figure 4-4, sorting the values in descending order on a per project basis. Again, this figure only depicts data from the 60 projects that reported labor hours. This type of figure could also help companies see how their projects compare both internally and externally.

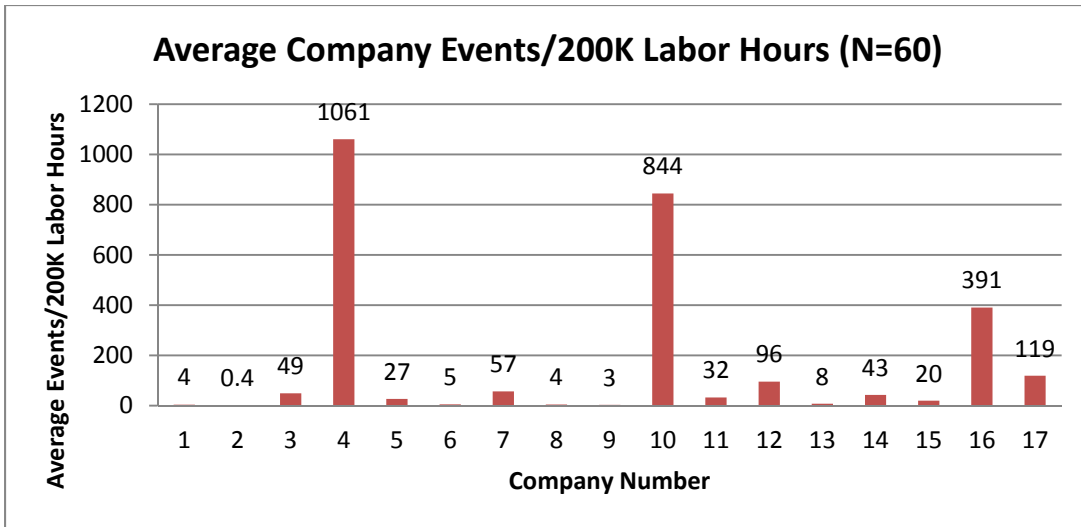


Figure 4-4. Average Normalized Events/200K Labor Hours Company Comparison

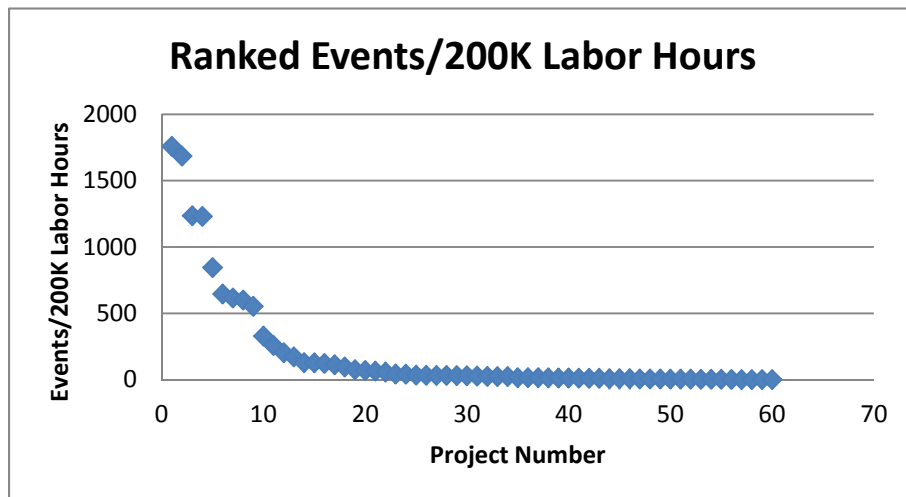


Figure 4-5. Descending Normalized Events/200K Labor Hours per Project

Tables 4-12 and 4-13 show the minimum, maximum, mean, median, and standard deviation of the normalized appraisal and events data for owners and contractors. These tables were constructed to illustrate the variation in the data set. After the activities and events were normalized per labor hours, the variation from project to project was reduced drastically, as seen

in the standard deviation section. Owners still have a large standard deviation for appraisals and defects and contractors, again, for appraisals.

Table 4-12. Owners Descriptive Statistics for Normalized Activities/Events per Labor Hour

	Appraisals	Findings	Variations	Defects	Failures
Minimum	0.0	0.0	0.0	0.0	0.0
Maximum	1619.4	466.7	88.9	1669.8	37.0
Mean	226.4	30.3	12.5	328.2	4.5
Median	43.7	1.9	3.7	9.7	0.0
Std. Dev.	374.9	100.9	22.3	567.3	10.3

Table 4-13. Contractors Descriptive Statistics for Normalized Activities/Events per Labor Hour

	Appraisals	Findings	Variations	Defects	Failures
Minimum	0.0	0.0	0.0	0.0	0.0
Maximum	11,685.9	496.6	623.7	148.5	72.8
Mean	754.6	32.7	47.4	17.8	2.5
Median	39.0	11.4	0.0	4.5	0.0
Std. Dev.	2005.7	79.6	152.7	33.1	11.6

4.4.3. Hypotheses

In order to test the validity of the data collection process and the quality metrics program that was developed, a total of five hypotheses were developed. Since this study was based on pilot data and the RT313 research is still ongoing, these hypotheses describe the relationship that might exist amongst the planned activities and unplanned events, as well as, the pyramid structure, and the research assumptions. Statistical hypothesis testing was then used to test these relationships. Table 4-14 shows the final hypothesis statements for the research along with a description of the desired outcome and the type of statistical testing that was used. Each hypothesis, and the results of its analysis, is discussed separately in further detail. Overall conclusions about the industry are not made as a result of this study as it is considered a preliminary analysis of a larger study still ongoing with CII RT313.

For purposes of the study, three different normalizers, project dollars, labor hours, and duration (in years), were used for all of the hypotheses. Ultimately, labor hours (at the level of 200,000 labor hours) were determined to be the normalizer that should be used because the individuals doing the work on the project are typically the cause of the quality event. Also, the amount of labor hours of a project is a good indicator of its magnitude and thus would allow the projects to be compared in similar terms. Recall that labor hours was also the measurement that was used in the denominator of the QPR formulation.

Table 4-14. Hypothesis Statements

No.	Hypothesis Statement	Description	Testing Method
1	The higher the number of prevention activities, the fewer the number of normalized quality events.	The number of normalized quality events per project should be negatively correlated with the number of normalized prevention activities per project.	Pearson Correlation Analysis
2	The higher the number of appraisal activities, the fewer the number of normalized quality events.	The number of normalized quality events per project should be negatively correlated with number of normalized appraisal activities per project.	Pearson Correlation Analysis
3	The higher the level self-reported project quality the fewer the number of normalized quality events.	The number of normalized quality events per project should be negatively correlated with the self-reported perception of quality by project.	Pearson Correlation Analysis
4	The higher the level of self-reported project quality the fewer the number of normalized quality events higher on the severity scale.	The average number of each normalized quality event per project should be increase as the level of the self-reported perception of quality decreases by project.	Two-sample mean hypothesis test
5	In a given project, the frequency of findings will be higher than variations, the frequency of variations will be higher than defects, and the frequency of defects will be higher than failure.	The proportion of quality events to total events per project should increase as the severity level decreases.	Two-sample mean hypothesis test

Hypothesis 1. The first hypothesis states that the more prevention activities per project, the less normalized quality events should occur. This kind of relationship would allow for a correlation analysis to be conducted as shown in Table 4-15. In order for this relationship to be true, a negative correlation should exist.

Table 4-15. Hypothesis 1 Correlation Analysis

Prevention to Normalized Events Correlation Coefficient			
	Events/\$	Events/Labor Hour	Events/Year
Owners	0.208	0.020	0.074
Contractors	0.134	0.202	0.156
All	0.124	-0.019	-0.017

As shown in the table above, the submitted projects did not show a strong correlation toward this hypothesized relationship. Because of this, the data was broken down by project size, consisting of mega, large, and small classifications to try and identify a subset of projects that could illustrate this hypothesis. As shown in Table 4-16, mega projects had a correlation coefficient of -0.100. This correlation coefficient is still not large enough to be considered strong.

Table 4-16. Hypothesis 1 Subgroups

Prevention Correlation in Project Size Subgroups		
Project Size	N	Correlation Coefficient
Mega (greater than \$1B)	41	-0.100
Large (between \$10M and \$1B)	12	-0.016
Small (less than \$10M)	7	0.805
Overall	60	-0.019

Hypothesis 2. The second hypothesis aimed to test the relationship between appraisal activities and quality events. Similar to Hypothesis 1, a negative correlation would be desired because of the idea that appraisal activities should reduce the number of quality events that occur. Table 4-17 shows that while there is more evidence to prove an inverse relationship between appraisals

and quality events, it is again, not significant. Notice that the test that includes the QPR (Events/Labor Hour) shows a slight negative correlation. This relationship between appraisals and the QPR is illustrated in Figure 4-6. Also notice that owner companies have weaker correlations in all normalizing categories, especially the strong positive correlation found when normalizing by years. This is consistent with the findings of the demographic study that found that owners less effective appraisal processes.

Table 4-17. Hypothesis 2 Correlation Analysis

Appraisals to Normalized Events Correlation Coefficient			
	Events/\$	Events/Labor Hour	Events/Year
Owners	-0.048	0.379	0.933
Contractors	-0.071	-0.169	0.024
All	-0.019	-0.067	0.121

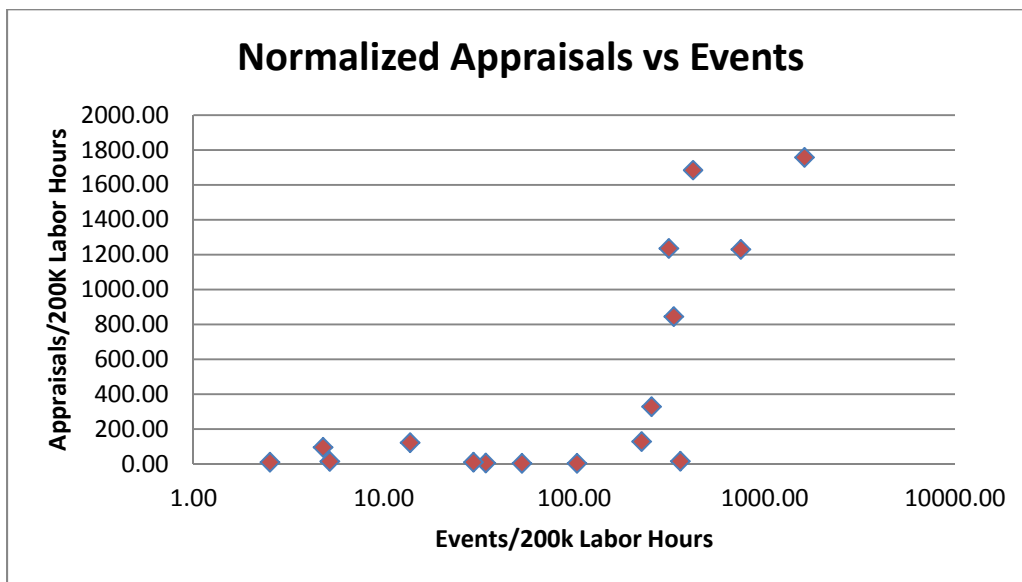


Figure 4-6. Hypothesis 2 Graph

Due to poor correlation results, the same analysis was done using subgroups based on project size. This particular subgroup of the data was chosen because the project size is thought to be a source of some of the variation in the data. In this analysis, large projects ended up having the

strongest correlation of appraisals to normalized quality events (-0.248), however, this is still considered weak.

Table 4-18. Hypothesis 2 with Subgroups

Appraisal Correlation in Project Size Subgroups		
Project Size	N	Correlation Coefficient
Mega (greater than \$1B)	41	-0.020
Large (between \$10M and \$1B)	12	-0.248
Small (less than \$10M)	7	0.035
Overall	60	-0.067

Hypothesis 3. This hypothesis tested the self-reported perception of quality rather than the structure of the pyramid. Namely, the higher the level of self-reported project quality, the fewer the number of normalized quality events. To test this hypothesis, a correlation coefficient was calculated between the values of the normalized quality events and the self-reported perception of quality. A strong negative correlation across all projects would indicate that the industry was able to judge the relative quality of a project without the quality pyramid being implemented into their company. Based on the results shown in Table 4-19, four tests had a negative correlation and five tests had a positive correlation. None of the correlation coefficients were strong enough to indicate a strong correlation.

Table 4-19. Hypothesis 3 Correlation Analysis

Quality Self Evaluation Correlation Coefficient			
	Events/\$	Events/Labor Hour	Events/Year
Owners	0.058	0.299	0.002
Contractors	-0.027	-0.351	-0.016
All	0.038	-0.012	0.003

Hypothesis 4. The fourth hypothesis aimed to test the relationship between the self-reported perceptions of quality to normalized quality events. More specifically, the self-reported

perceptions of quality would be split into subgroups comparing the normalized quality events for those projects reports a “strongly agree” and “agree” versus projects reporting a “strongly disagree” and a “disagree”. A two-sample mean hypothesis test would be conducted to determine if the average values of the two subgroups were statistically significant. As shown in Table 4-8 earlier in the chapter, there was a total of 70 “agree” and “strongly agree” project while there was only five “disagree” and “strongly disagree” projects. This very small sample of “disagree” type projects did not allow this hypothesis to be tested.

Hypothesis 5. The fifth and final hypothesis aimed to test the structure of the unplanned events portion of the pyramid. In particular, it states that the proportion of findings to total events will be higher than the proportion of variations to total events, the proportion of variations to total events will be higher than proportion of defects, and the proportion of defects to total events will be higher than the proportion of failures.

This hypothesis was tested using a two-sample mean hypothesis test to test for a significant difference between the two samples. Table 4-20 shows the average values for the proportions of each quality event for both owners and contractors. Table 4-21 and Table 4-22 show the p-values from each test for owner and contractor projects, respectively. Note that while the p-value for variations versus defect is significant for both owners and contractors, the defects were the large sample group. Therefore, the p-value that would show whether there was a significant difference between variations and defects, and the amount of defects was less than variations would be found by the formula, $1-p$. Because of this, these tests actually fail to statistically show that the proportions of variation to total event are greater than the proportion of defects to total events for both owners and contractors.

Table 4-20. Average Proportions of Events

Event	Owners	Contractors
Findings	0.17	0.56
Variations	0.11	0.13
Defects	0.68	0.28
Failures	0.03	0.03

Table 4-21. Hypothesis 5 Owners Analysis

Owners (N=46)	
Comparison	P-Value
Findings vs. Variations	0.107
Variation vs. Defects	0.999
Defects vs. Failures	<0.001*

*Note 1: Let * denote a significant difference using a 95% confidence interval.*

Table 4-22. Hypothesis 5 Contractors Analysis

Contractors (N=56)	
Comparison	P-Value
Findings vs. Variations	<0.001*
Variation vs. Defects	0.992
Defects vs. Failures	<0.001*

*Note 2: Let * denote a significant difference using a 95% confidence interval.*

4.4.4. Data Analysis Conclusion

While many of the analyses that were developed did not have conclusive results, the data collected during this study is considered pilot data and therefore it would be premature to draw conclusions from this stage of the work alone. Moreover, there are many opportunities to introduce variation in this data. For example, the quality pyramid program had to be retroactively applied to past projects within the participating companies. This means that each company had to consider the definitions of the quality events and apply them to any existing metrics or data that they saw fit.

Therefore this implementation process was dependent not only upon the individual's ability to correctly map each quality event to the appropriate set of data within their existing QMS but also to the maturity that QMS and its amount of available data. It is important to note that variations in maturity of a company's QMS could lead to significant variations in data reporting. For this reason, the data summaries and analyses shown in this chapter were not done to infer final conclusions, but to simply show the kind of tests that could be done once this program has been uniformly implemented into companies throughout the capital facilities delivery industry.

Next in Chapter 5, the conclusions from the pilot data collection study and future work is discussed.

5. Conclusions and Future Work

5.1. Summary of research goals and objectives

The purpose of this research was to create a standard for industry-wide quality metrics. The work completed for this master's thesis is subset of this larger effort being conducted by RT313 and funded by the Construction Industry Institute. Specifically, the primary objective of the research aimed to develop a set of quality metrics that would allow both owners and contractors to:

- Replicate as much as possible the characteristics present in today's accepted safety metrics;
- Define what a quality failure is;
- Identify areas for improvement;
- Establish a common understanding (common language relative to quality similar to the common language utilized in measuring safety performance);
- Benchmark current performance and identify trends both internally and across the industry;
- Support ability to initiate pro-active actions to assure maintaining or improving quality in lieu of continually reacting to quality issues;
- Identify best practices and provide input that allows for work process improvements that demonstrate value to the business by positively impacting the "bottom line" (quantifying the cost of failure to support maintaining quality and understanding the impact of non-conformance); and
- Demonstrate quality performance relative to the industry to internal and external customers and use the performance indicated by the metrics to differentiate themselves.

The secondary objective of the research aimed to support a culture change where quality issues are identified, reported and measured similar to the changes that occurred in awareness of safety issues resulting in the current safety conscious culture that encourages and rewards continuing

improvement (change behaviors). This chapter will summarize the outcomes of the research and address the whether each objective was achieved.

5.2. Summary of Research Outcomes

5.2.1. Outcomes from the Literature Review

The literature review first surveyed the general BOK relating to quality and quality metrics. This study found that metrics should:

- Be few in number;
- Encompass all quality concerns in the scope of work;
- Be clearly defined and easy to understand;
- Drive improvement;
- Be easy to collect.

Next, the literature review surveyed the BOK both inside and outside of the capital facilities delivery industry. Based on the review of literature within the industry, no standardized quality metric is currently being used uniformly, across the industry. The majority of the existing metrics focus on efficiencies of cost and schedule rather than the quality of the deliverable itself.

Interestingly, several associations outside of the capital facilities delivery industry are actively developing quality metrics specific to an industry such as the FDA for the pharmaceutical manufacturing industry or general, such as APQC. This supports the need for the development of a pre-determined standardized common quality metrics framework for the capital facilities delivery industry.

5.2.2. Observations from the Data Collection

During the data collection process, the research team made numerous observations related to the potential impact that the program could have on the industry. The bulleted list below provides a

summary of these observations that were supported from lengthy discussions and face validity of the SMEs from the data.

- Cultural change is necessary to ensure that quality events are accurately reported, similar to safety metrics.
- SMEs learned that it is harder to input data on closed projects. It is difficult to obtain data for project that have been handed-off and thus, the quality of that data is questionable. Therefore, a new quality metrics framework is best implemented for new or in-process projects.
- Some companies have indicated that no quality events are reported to them after hand-off. This could cause a lower reporting of failures than what actually occurs. Moreover, the absence of quality events data does not imply that there were no quality issues, rather, it indicates that there was a lack of reporting quality events.
- There is an inherent bias in the data collection since it is all from CII members.
- Owner companies have shown a resistance toward reporting labor hours while contractors resist reporting project cost.
- While labor hours have been the recommended normalizing factor, the team recognizes that the amount of labor hours required could be dependent on the country. Example: Labor hours in China may be 10 times higher for the same work done in the U.S. This inherent characteristic of the industry could make it difficult to compare projects to projects or company to company due to the variation in the normalizing factor.
- A potential for duplication of reporting is possible for quality metrics. For example, in safety, owners report all safety data for the project, while the contractors only report data

for their own company. This results in contractor safety data being reported twice.

Similar results could occur under this quality metrics program.

- Open projects have a potential to skew the data because they will likely have quality event occur after the data has been reported.
- The amount of data reported could be dependent on the maturity of a company's QMS. Therefore, comparing companies with by QMS maturity level could reduce variation in the QPR.
- Company's individual standards of reporting could also cause variation. For example, companies that are very strict in their design standards may have more defects than variations due to their inability to deviate from requirements.

Many of these observations of the potential impacts on the industry point out the complexity of creating a standard quality metric for the entire capital facilities delivery industry. They also illustrate that many aspects of the industry have a potential to reduce variation of the reporting of the quality metrics and the calculation of the QPR. In reference to the overall project goals and objectives, the research team concluded that:

- The Quality Pyramid is a novel and visually appealing approach to measuring quality and can be used across the project lifecycle to measure quality performance.
- The quality framework is easy to use and applicable within the industry across different project types and sizes for owners and contractors.
- The quality metrics use a common language that enables implementation within an organization.
- The quality metrics can drive improvement within organizations internally and allow for external benchmarking against peers or the industry as a whole.

- The Quality Performance Rate (QPR) is an indicator of company and project level quality allowing for benchmarking and continuous improvement of a QMS.

5.2.3. Outcomes of the Data Analysis

The data analysis consisted of summarizing demographic type information as well as, summaries of the quality activities and events, and hypothesis testing. The following are a list of key findings from the data analysis:

- 35 projects were successfully reported from CII members outside of RT313, showing that the data collection protocol is clearly defined and easy to understand.
- The majority of reported projects (74 out of 102) are from heavy industrial.
- Contractors perform more prevention activities than owners.
- Owners tend to not report labor hours.
- Contractors tend to not report total project cost.
- 86 out of the total 102 projects reported “somewhat agree” or higher on their self-reported perception of project quality. This shows that Quality SMEs are not able to correctly identify a projects level of quality with current metrics and QMSs.
- Owners have more normalized defects than findings while contractors have more normalized variations and defects than findings.
- Data is best normalized using labor hours.
- While owners conduct roughly one third of the appraisals that contractors do, owners report an average of 29.5 normalized findings per project versus 7.8 for the contractors.
- For owners, there is a high variability in appraisals, findings, and defects per project while contractors had a high variability in appraisals per project.

- Submitted data using past projects show a high variability. This could be due to a number of things:
 - There could be a better normalizer that has not yet been discovered.
 - QMSs vary in maturity.
 - User error when retroactively applying the pyramid to current QMS.
 - Variability in company/industry standards of reporting.
- This variability led to inconclusive hypothesis testing.

5.3. Conclusions

This research effort demonstrates that companies are able to align project data and report against the events in the quality pyramid within their current systems. Reporting quality data in this manner allows companies to produce a QPR similar to the OSHA Rate in order to assess individual project performance. This study is meant to show interesting outcomes of the pilot data collection study as well as the types of analyses that could be done once the full quality metrics program has been further developed and approved for implementation by CII. No concrete conclusions are meant to be drawn from the results of the study. While quality issues can never be totally eliminated, focusing on eliminating the causes of defects and failures in the project systems pays dividends in successful project completion.

5.4. Future Work

Future work for this research effort has been identified as it relates to project enhancement and for the successful implementation and adoption.

Future work directly relating to project enhancement might include:

- Further develop the prevention and appraisal activities.

- Further develop of the severity model. How will the metric include the severity of each event?
- Expand the data set. Possibly work with fewer companies to get more detailed project data.
- Work to better understand how respondents reported their data. For example, how did they classify each event with their existing company data? This process could uncover some explanations to the variations seen in the data.
- As owners collect data, is there a correlation between the QPR and those owners that use their own internal engineering versus those owners who use an EPC?
- Continue to investigate statistical hypothesis testing to validate the pyramid.
- Investigate potential cause of the variation in the data.
 - Ask companies to report on the maturity of their QMS at the time project data is submitted.
 - Ask companies to report on their standard of reporting of quality events data at the time project data is submitted.
 - Verify that companies have correctly assigned the quality events to their QMS.
- Investigate other normalizers that have not yet been considered.
- Determine what is considered an outlier and if it should be eliminated from the data to conduct further, more refined, testing.

Future work relating to implementation and adoption include:

- Get buy-in from CII to be data repository.
- Obtain data for projects after the program has been implemented to see if the hypotheses uphold.

- CII-wide adoption.
- Industry-wide adoption.
- Possibly develop a data collection tool that would allow quality events to be logged in real-time or as they occur. This would help eliminate having to report data on closed projects.
- In general, utilizing labor hours as the normalizing factor works well, but fabricated equipment companies, for example, may find it hard to report. Work needs to be done on handling these types of specific cases to ensure uniform adoption.
- Ideally, project data will be collected internally and in real-time throughout the lifecycle of the project. Once the project is completed, the finalized data from all quality events would be reported to CII. This would eliminate the difficulty of reporting data on closed projects, as well as, the possibility of data from open projects skewing the data sample.

References

- ASQ. (2013). "ASQ - The Global Voice of Quality.": www.asq.org.
- ASQ. (2013). "The Global State of Quality Research Discoveries 2013.": (<http://asq.org/global-state-of-quality/reports.aspx>)
- Bird, Frank E., Germain George L.(1996). Loss Control Management: Practical Loss Control Leadership, Revised Edition, Det Norske Veritas (U.S.A.), Inc, Figure 1-3.
- Buttell, Phil; Hendler, Robert; Daley, Jennifer; (2008). "Quality in Healthcare: Concepts and Practice." *The Business of Healthcare*, 68-90.
- Cha, Hee Sung; Kim, Chan Kyu (2010). "Quantitative Approach for Project Performance Measurement on Building Construction in South Korea." *KSCE Journal of Civil Engineering*, 1319-1328.
- Costa, Dayana B.; Formoso, Carlos T.; Kagioglou, Michail; Alarcon, Luis F. (2004). "Performance Measurement Systems for Benchmarking in the Construction Industry." *Proc. 12th IGLC Conf. LO-School*, Elsmore, Denmark, 451-463.
- Crosby, P. B. (1962). Quality management: the real thing.: <http://www.wppl.org/wphistory/philipcrosby/QualityManagementTheRealThing.pdf>
- Damle, Subhash. (Forthcoming). "CII RT313 Implementation Resource for Creating Quality Metrics for the Capital Facilities Delivery Industry"
- Deming, W. E. (1986). Out of the Crisis, 1986. *Cambridge, Mass.: Massachusetts Institute of Technology Center for Advanced Engineering Study*. xiii, 507.
- DOE (2009). "Human Performance Improvement Handbook.":<http://www.hss.energy.gov/nuclearsafety/ns/techstds/>
- FDA. "U.S. Food and Drug Administration." *Drug Compliance Programs*. U.S. Food and Drug Administration, n.d. Web. 17 Nov. 2013.
<<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/ucm252671.htm>>.

- Heinrich, H. W. (1959). *Industrial Accident Prevention*, 4th Edition. New York: McGraw-Hill Book Company, Inc.
- Juran, J.M.; Gryna Jr., Frank M. (1980). *Quality Planning and Analysis*, Second Ed., US: McGraw-Hill, Inc.
- Kaplan, Robert S. and Norton, David P. (1992). "The Balanced Scorecard – Measures That Drive Performance." *Harvard Business Review*. 71-79.
- Lamb, K., Needy, K. L., Ries, R., & Mahjan, V. (2014). Project quality metrics inside and outside the capital facilities delivery industry. 2014 American Society for Engineering Management Proceedings. Virginia Beach, VA.
- Masimore, Lyle. (2007). "Proving the Value of Safety." *SAFETY*. 11.
http://literature.rockwellautomation.com/idc/groups/literature/documents/wp/safety-wp004_-en-p.pdf
- Mendivil, Steve, et al. "PARENTERAL DRUG ASSOCIATION POINTS TO CONSIDER: Pharmaceutical Quality Metrics Updated September 2014." *PDA Journal of Pharmaceutical Science and Technology* 68.5 (2014): 535-545.
- Needy, Kim LaScola. (2013). "CII RT313 – October Meeting Minutes."
- Needy, Kim LaScola; Ries, Robert. (2013). "CII RT313 Proposal – Creating Standards for Industry-wide Quality Metrics."
- Needy, Kim LaScola, Ries, Robert, Lamb, Kelsey. (Forthcoming). "CII RT313 Research Report – Creating Quality Metrics for the Capital Facilities Delivery Industry"
- Parental Drug Association (PDA) (2013). "Pharmaceutical Quality Metrics":
<http://www.pda.org/pdf-1/PDA-Pharmaceutical-Quality-Metrics.aspx>
- Sklamberg, Howard R. (2013). "Evolution in FDA's Approach to Pharmaceutical Quality." (<http://www.fda.gov/downloads/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCA/SignificantAmendmentstotheFDCA/FDASIA/UCM363523.pdf> , Accessed September 10th, 2013).

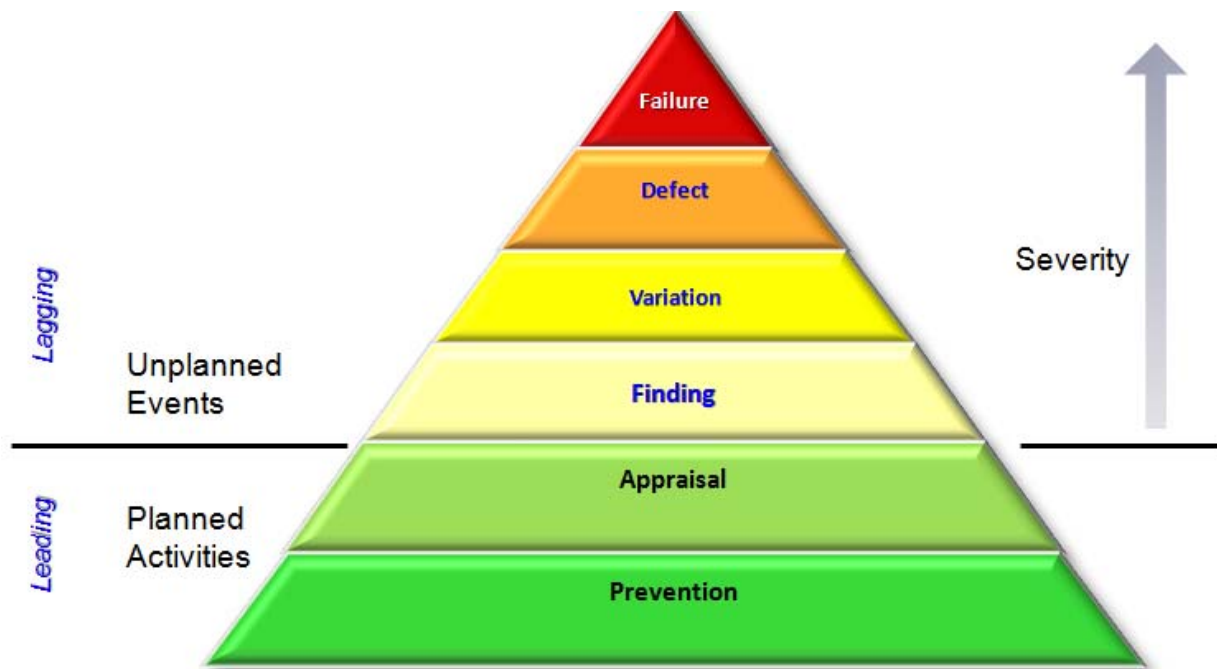
Taguchi, G. (1986). Introduction to quality engineering: designing quality into products and processes.

Training Resources and Data Exchange (TRADE), U.S Department of Energy - Defense Programs. (1995). "How to Measure Performance: A Handbook of Techniques and Tools." 53-58.

Thor, Carl G. (2006). "Using a Family of Measures to Assess Organizational Performance." *Global Business and Organizational Excellence*, 14(3), 111-131.

Quality Pyramid Illustrations

CII – RT313



Prevention

Definition

Activities for preventing nonconformance in products or services through the use of established procedures, instructions, tools and other applicable documents.

Description

Procedures, processes, tools and resources employed to continuously lower the risk of Quality nonconformance. Think of this activity as part of the cost of Quality.

Examples

- Verify that Quality Procedures exist, for example:
 - Deliverable Verification
 - Corrective action / preventive action (CAPA) process
 - Equipment QA Inspection procedures
 - Management Quality review process
 - Quality Assurance Manuals
- A Project Execution Plan exists which includes, as a minimum, the Project Quality Plan. The Quality Plan should address all applicable phases of the project.
- Quality infrastructure exists with an organization on project, office and overall organization levels. This includes the necessary people resources and tools to ensure the quality process is in place
- Owner's Quality Roles and Responsibilities are well documented when work is contracted
- Corporate Quality Metrics procedure exists and/or dashboard
- Contract Partners Quality Programs meet all expectations and requirements
- Quality training activities are in place

Appraisal

Definition

Planned and documented activities performed to determine the adequacy, effectiveness, and compliance with, established procedures, resources, instructions and/or drawings and other applicable documents.

Description

An appraisal is conducted to assure process and/or work product compliance by identifying the key measurements and tolerances related to the process or product.

Examples

- **Audits** – A formal, systematic and objective activity conducted by a qualified individual to verify compliance to specified requirements. This could include evaluation of an overall management system, e.g. execution plan, standards, procedures, etc., or a discipline work product such as drawings, data sheets, calculations, etc.
- **Surveillance** – The systematic monitoring of activities (or behaviors) to assure that processes are properly implemented and requirements are being met. This is normally accomplished by the use of assessment and verification activities.
- **Testing** – Destructive or non-destructive examination to determine the characteristics of a system or component or material. This includes written examinations of personnel qualifications. (include examples of an assessment)
- **Assessments** – An investigation and analysis of work management systems, surveillance data, procedures and processes to determine suitability, effectiveness, and compliance with specified requirements. (include examples of an assessment)
- **Verifications** – A quality control activity to confirm that a product meets requirements. For example, this could be an engineering-discipline deliverable, a pump, a pressure vessel or an exchanger in a fabrication shop, or a field activity such as concrete placement, equipment setting, structural erection, piping installation, and similar activities. This includes positive material identification.

Finding

Definition

An outcome of an “Appraisal” that provides documented determination that either the compliance with or implementation of established procedures, resources, instructions or other applicable documents is inadequate. A Finding requires a documented response as to corrective and/or preventive actions with follow-up to assure implementation of stated actions.

Description

There is a process but there is a deficiency in how it is executed.

Examples

- A procedure that did not accomplish the intended function.
- A deficiency in the execution of required processes such as: inspections, audits, tests, checklists, protocols and documentation
- Execution of activities/processes by un-qualified personnel/organizations.
- Did not investigate and rectify known management system weaknesses.
- Did not implement required corrective/preventative actions.

Variation

Definition

A non-conforming item (system, structure, component or service) that has been determined through investigation, examination or evaluation of objective evidence can be accepted as is and may change the requirement.

Description

Acceptable *Variations* may take the form of documented and communicated design deviations or specification variances in which the results of acceptance do not compromise the intended form, function or purpose of the item. This is a nonconformance to a specification or design that can be “accepted as is” since it will not lead to a defect or a failure.

Examples

- During inspection of fabricated products, a tolerance violation was identified. Based on an engineering evaluation, the out-of-tolerance condition of the fabrication was determined to be acceptable. The resolution of the nonconformance resulted in a “use-as-is” disposition for the identified condition.
- During installation of the piping spool that mates with a heat exchanger nozzle, it was determined that the nozzle flange was 150 Lb. Class. The engineering specification and equipment supplier documents indicated the flange to be 300 lb. class. Based upon engineering evaluation, the 150 lb. class flange was determined to be acceptable for the system pressure and temperature conditions. The resolution of the nonconformance resulted in a “use-as-is” disposition for the identified condition.
- During the final walk down of a steam system prior to final pressure test, it was discovered that the relief valve discharge line was not supported as indicated on the engineer’s drawings. Based upon engineering evaluation, the as-installed support system was determined to be acceptable for the system conditions. The resolution of the nonconformance resulted in a “use-as-is” disposition for the identified condition.

Defect

Definition

A nonconforming item (system, structure, component or service) that has not fulfilled an intended requirement or reasonable expectation for use and requires correction.

Description

The definition of what constitutes a defect may vary according to the point in the quality chain at which it is discovered. A defect is corrected prior to transfer to operations/owner/end user.

Examples

- A weld does not pass inspection before it is turned over to operations and cannot be accepted as is.
- A supplier provides fasteners that were not properly heat treated and cannot be accepted as is.
- Commissioning activities reveal a system or component which does not work or does not work adequately and cannot be accepted as is.
- Modifications to an installed piping system results in loss of the required material traceability information. The installed piping cannot be accepted as is.

Failure

Definition

The inability of a system, structure, component or service to perform its required function on demand. Not detected or detectable within the quality program before utilization for intended purpose.

Description

Failure is an end user disruptive event where the system, structure, component or service does not meet agreed requirements. The event occurs after handover to operations/owner/end user.

Examples

- The facility does not run at design capacity
- Weld fails after turnover to operations causing the plant to shutdown
- A pump fails to perform at the design capacity
- If a defect shows up after handover it is a failure.

Appendix B: Confidentiality Agreement

Researchers: Robert Ries and Kim Needy

You are being asked to participate in a research study.

Please read the following before you agree to take part in this study.

Dear CII Member:

CII RT 313 – Creating Standards for Industry-Wide Quality Metrics is conducting a data collection effort to identify and define a capital facilities delivery industry set of quality metrics that can be used effectively throughout a project lifecycle to assess, analyze, measure and benchmark project performance.

Data is being collected using a Qualtrics survey hosted at the University of Florida. All study data will be collected through an online survey-collection program called Qualtrics. Qualtrics is a secure site with SAS 70 certification for rigorous privacy standards. Any data that you provide through this program will be encrypted for security purposes using Secure Socket Layers (SSL). Only the study investigators will have access to the data on Qualtrics. To protect your privacy, all participants' IP addresses will be masked by Qualtrics and will be unavailable to, and unidentifiable by, investigators or others. Qualtrics' privacy policy can be obtained at <http://www.qualtrics.com/privacy-statement>.

The survey contains instructions, a place to record company information, and an area to record project information including quality events.

Data collected will be stored by the academic research team at Qualtrics. Only the academic research team and designated CII staff members will have access to the data. Researchers will follow the Construction Industry Institute (CII) Data and Confidentiality Guidelines.

Participation in this research is strictly voluntary. If at any time you have a concern, feel free to contact Dr. Kim Needy, University of Arkansas, kneedy@uark.edu, 479-575-4401 or Dr. Robert Ries, University of Florida, rries@dcp.ufl.edu, 352-273-1155. Thank you for your participation.

You may contact the Institutional Review Board at the University of Florida Health Science Center at 352-273-9600 if you have questions about your rights as a research subject or what to do if you are injured.

You may choose not to be in this study or you may quit being in the study at any time and there will be no penalty and no loss of any benefits you are entitled to.

Information collected about you will be stored in locked filing cabinets or in computers with security passwords. Only certain people have the legal right to review these research records, and they will protect the secrecy (confidentiality) of these records as much as the law allows. These people include the researchers for this study, certain University of Florida

officials, the hospital or clinic (if any) involved in this research, and the Institutional Review Board (IRB; an IRB is a group of people who are responsible for looking after the rights and welfare of people taking part in research). Otherwise your research records will not be released without your permission unless required by law or a court order.

Researchers will take appropriate steps to protect any information they collect about you. However there is a slight risk that information about you could be revealed inappropriately or accidentally. Depending on the nature of the information such a release could upset or embarrass you, or possibly even affect your insurability or employability.

If the results of this research are published or presented at scientific meetings, your identity will not be disclosed.

Appendix C: Survey Instructions

Before Starting

1. Read the definitions for the following quality events: Finding, Variation, Defect, and Failure located in the *Instruction Worksheet* on the Excel tool that is being used for this research project. Map these quality events with the corresponding name(s) used at your own company. Note: you may have more than one term used to define each quality event.
2. For each event, identify where this data resides within your own quality management system. This will be helpful for locating quality event information needed to collect data on various projects.

Utilizing the Quality Metric Tool

Company Data Tab

1. Enter required information highlighted in green for your company.

Template Worksheet

The template provides a layout for data collection. Data should not be entered into the template; rather this worksheet can be copied for data to be collected on various projects. Projects could be a closed ongoing. If data is being reported for more than one project, a separate worksheet should be created by copying and pasting the template tab for each new project. Label these worksheets, Project 1, Project 2, Project 3, etc.

Cells in the Excel spreadsheet that are flagged in red contain helpful information.

Data may be reported in summary by entering the total number of findings, variations, defects, or failures into the *Project Summary Table* directly (**not preferred**), or by entering each event separately into the *Event Log Table* (**preferred**). If data is entered directly into the *Event Log Table*, the *Project Summary Table* will be automatically calculated for you.

Fields highlighted in green are required, if available.

Steps for Completing the Template Worksheet:

1. Create a new project worksheet by copying and pasting the template worksheet and retitling it to reflect the new project name such as, Project 1, Project 2, Project 3, etc.
2. At the top of the worksheet enter the *Project Information* data. Note: Some companies may not be able to report both cost and labor hours data, but if both data are available, please report.
3. Enter quality event data by either using the *Project Summary Table* or the *Event Log Table*.

- a. If using the *Project Summary Table* (**not preferred**), enter the total number of Findings, Variations, Defect, and Failures by typing over the formulas.
- b. If using the *Event Log Table* (**preferred**), enter each quality event in a separate row. Include as much information as possible as this will help with data validation. Note: For this option, the *Project Summary Table* will be automatically calculated for you.

Repeat steps 1-3 for the next project.

Appendix D: Qualtrics Survey

Construction Industry Institute (CII)

RT-313: Creating Standards for Industry-wide Quality Metrics

Q1 Your name (required)

Q2 Email (required)

Q3 Organization Name

Your Position / Title in Your Organization

Business Phone

Q4 Company Type

Owner

Contractor

Q5 Industry Group for the Project

Heavy Industry

Light Industry

Building

Infrastructure

Q6 Project / Contract Size

Small (less than \$10M)

Large (between \$10M and \$1B)

Mega (greater than \$1B)

Q7 Project Status

Open

Closed

Q8 Location Country

Q9 Project / Contract Start Date

Q10 Project / Contract End Date or Anticipated End Date

Q11 Total Project and Contract Cost and Work Hours to Date (required)

Either Total Cost or Total Work Hours must be entered

Entering both Total Cost and Total Work Hours is preferred

If the project is closed, please enter the total cost (in US\$) based on the company's scope of work. If the project is open, please enter the cost to date (in US\$) based on the company's scope of work. If the project is closed, please enter the total work hours based on the company's scope of work. If the project is open, please enter the work hours to date based on the company's scope of work.

Q12 Project Evaluation (required)

Please select the answer that best reflects your agreement with the following question:

This is a high quality project.

- Strongly Disagree
- Disagree
- Somewhat Disagree
- Neutral
- Somewhat Agree
- Agree
- Strongly Agree

Q13 Prevention Activities

Please check each of the following types of activities that were or are currently being performed on the project:

- Quality procedures are implemented
- Corporate quality metrics and/or procedures are implemented at the project level
- A quality infrastructure is implemented within the project
- Quality roles and responsibilities are documented
- Contract partners have implemented quality programs
- Quality training activities are in place and have been executed
- Scope and criteria management systems are implemented
- A project execution plan includes a project quality plan for the project
- Stakeholder alignment has been achieved
- None

Q14 Please enter total number of appraisal, finding, variation, defect, and failure events below.

_____ Appraisal
_____ Finding
_____ Variation
_____ Defect
_____ Failure

Appendix E: IRB Letter



Office of Research Compliance
Institutional Review Board

June 10, 2014

MEMORANDUM

TO: Kim LaScola Needy
Kelsey Lamb
Robert Ries
Vishal Mahajan

FROM: Ro Windwalker
IRB Coordinator

RE: New Protocol Submission

IRB Protocol #: 14-05-755

Protocol Title: *RT 313 - Creating Standards for Industry-Wide Quality Metrics*

In reference to the request for IRB approval of your project titled *RT 313 - Creating Standards for Industry-Wide Quality Metrics*, the IRB is not authorized to oversee and approve such research. This protocol does not meet the definition of research involving human subjects in the federal regulations. (See the citation below.) You are free to conduct your research without IRB approval.

45 CFR 46.102 (f)

(f) Human subject means a living individual about whom an investigator (whether professional or student) conducting research obtains

- (1) Data through intervention or interaction with the individual, or
- (2) Identifiable private information.

If you have any questions do not hesitate to contact this office.

210 Administration Building • 1 University of Arkansas • Fayetteville, AR 72701
Voice (479) 575-2208 • Fax (479) 575-3846 • Email irb@uark.edu

The University of Arkansas is an equal opportunity/affirmative action institution.