JOURNAL OF MANAGEMENT AND ENGINEERING INTEGRATION

Scope: The Journal of Management and Engineering Integration (JMEI) is a double-blind refereed journal dedicated to exploring the nexus of management and engineering issues of the day. JMEI publishes two issues per year, one in the Summer and another in Winter. The Journal's scope is to provide a forum where engineering and management professionals can share and exchange their ideas for the collaboration and integration of Management and Engineering research and publications. The journal will aim on targeting publications and research that emphasizes the integrative nature of business, management, computers and engineering within a global context.

EDITORIAL REVIEW BOARD

REVIEWERS

The Journal Editorial Team would like to thank the reviewers for their time and effort. The comments that we received were very constructive and detailed, and help us to continue to produce a consistently top-quality journal. Your participation is very important in the success of providing a distinguished outlet for original valuable articles. Again I would like to thank you all for your assistance in the review process. Below are the reviewers for the Summer 2019 issue.

Ron Barrett- Gonzalez, Ph.D.
Editor-in-Chief

Ali Ahmad	Jarrett Kotrozo	R. Radharamanan
Brooke Wheeler	Jason T. Black	Rian Mehta
Deborah S. Carstens	John Wang	Rick Wallace
Dennis Ridley	Junyong Ahn	Roelof deVries
Dia Ali	Kellie Schneider	Ronald Barrett
F. Michael Ayokanmbi	Khalid Bachkar	Ronald Shehane
Feng Zhou	Lauren Schumacher	Sandy Furterer
Gamal Weheba	Mary McShane Vaughn	Scott Swain
Gordon Arbogast	Mohammed Ali	Stephen Allen
Holger Mauch	Nabin Sapkota	Stephen Frempong
Huitian Lu	Narasimha Nagaiah (Raju)	Tamer Mohamed
Isaac Heim	Paul Nugent	Tristan Davison

TABLE OF CONTENTS

Smart Sensor Interface Model Using IoT for Pipeline Integrity Cathodic Protection

Mohammed Algarni[1]
Mohammed Zwawi[1]
[1]*Mechanical Engineering Department, Faculty of Engineering*
King Abdulaziz University, Saudi Arabia
malgarni1@kau.edu.sa; mzwawi@kau.edu.sa

Abstract

This study proposes a design of a smart sensor interface for monitoring Cathodic Protection (CP) in an Internet of Things (IoT) environment. The corrosion sensor is used to monitor voltage potential from a sacrificial anode to an underground pipe or pipelines. The past design of monitoring CP was via wired networking (for data transfer) or by onsite inspection. Although such inspection offers a very reliable monitoring method, it is expensive and inconvenient. This research proposes a new, advanced and straightforward method to monitor CP, for field engineers, by the advancement of IoT intelligence to provide suggestions and make online-orders for consumable parts. This research proposes a solution by the novel design of monitoring corrosion via CP using a smart sensor interface model in an IoT environment and smart-phone applications. The results of this system design show good monitoring performance and excellent analytical potential.

1. Introduction

This research investigates opportunities for integrating two different systems — mechanical engineering and information technology (IT) — into one integrated system. In the last decade, non-internet driven electronics and software-controlled mechanical systems have been investigated. However, these individual software solutions were not particularly efficient in communicating to the system owner or manufacturer (Fadell et al., 2017). Therefore, the idea of connecting traditional mechanical systems to the Internet to offer significant amounts of data to both system owner and manufacturer is highly promising. Accordingly, a new technology wave is glowing in the horizon, referred to as the “Internet of Things” (IoT), with considerable opportunities to change how mechanical systems will be designed and developed (Lee & Lee, 2015). Today, most mechanical systems such as elevators, air-conditioning and heating systems, plumbing, escalators, lighting, and power systems in large buildings are managed through a computer-based control system termed building a management system (BMS) that controls and monitors the mechanical systems (Minoli, Sohraby, & Occhiogrosso, 2017). Similarly, mechanical systems in petrochemical plants and refineries such as cathodic protection systems, pressure and safety valves, temperature and pressure sensors, piping and pressure vessels degradation inspection data, pumping and heat exchangers performance assessment, NDT, and safety and alarm systems are not all controlled by a computer-based control system. On the other hand, other existing systems have their own isolated software packages such as RIS, SAP, STID, and ARIS that work individually with no interaction with other systems (Ratnayake & Kusumawardhani, 2013). However, IoT offers to overcome all challenges that mechanical systems have raised throughout the past years. Integrating mechanical systems to the IoT will offer to collect, process, save data, and send recommendations and requests to the system owner. This research proposal will focus on a new application of integrating IoT to a Cathodic Protection system (CP).

2. Literature review

In industrial systems, wireless telecommunications have been rapidly advanced through IoT. It is expected that integrating IoT with the industrial telecommunication will increase profits in many applications (i.e. car factories and petrochemical plants) (Chi, Yan, Zhang, Pang, & Da Xu, 2014). Many other fields today are extensively implementing IoT devices into their existing systems. Firefighters have been implementing IoT into their field operation devices for rescue and research purposes. Similarly, medical implants have been implementing IoT for gathering and monitoring data. Likewise, other fields have been following the same approach such as with respect to farm animal biochips, city water supply and distribution, energy harvesting, crowd monitoring and automobile inherent sensors (Porkodi & Bhuvaneswari, 2014). An example of a design study in the IoT environment is proposed for the water quality of municipal corporations' monitoring systems. The study applied the smart city concept by designing a smart sensor interface to monitor water quality parameters such as temperature, water hardness, oxygen concentration, pH, and turbidity for cities using natural water resources such as rivers (Salunke & Kate, 2017).

3. Cathodic protection technique

Corrosion causes dangerous and expensive damages to oil and gas metallic pipelines, highway bridges, energy industries, storage tanks, ports, reinforced concrete, etc. According to a study by the National Association of Corrosion Engineers (NACE), the US annual direct cost of corrosion is estimated to be $276 billion. However, indirect costs such as labor activity, losses in productivity, delays, litigation, and factory failures have been estimated to total $552 billion annually (Koch, Brongers, Thompson, Virmani, & Payer, 2002). Nevertheless, after many decades of extensively studying corrosion, engineers have invented many approaches to help control it, but with a cost. Corrosion control approaches include CP, lining and coatings, corrosion inhibitors, and materials selection. Each approach can mitigate the corrosion process in different ways but cannot halt it. Corrosion (or rust) is a natural electrochemical deterioration process of materials that occurs in aqueous environments under ambient temperature. The process of corrosion (figure 1 *left*) involves electron removal of a metal, as in Equ. (1) (known as anodic reactions) and electron consumption at the same local metal in Eqs. (2) & (3) (known as cathodic reactions). In other words, the electrons flow from the anode part of the circuit to the cathode part, as shown in figure 1 (*right*). To ensure that this electrochemical process lasts, four components are required: anode, cathode, metallic path that connects the anode to the cathode, and an electrically conductive electrolyte for the electrons to flow, as illustrated in figure 1 *left*.

$$Fe \rightarrow Fe^{++} + 2e^- \quad (1)$$

$$O_2 + 2H_2O + 4e^- \rightarrow 4OH^- \quad (2)$$

$$2H_2O + 2e^- \rightarrow H_2 + 2OH^- \quad (3)$$

The nature of the electrochemical process of corrosion provides an opportunity to detect corrosion by using a voltmeter to monitor the delta-voltage between two different metals. In a case of having only one metal, such as underground pipelines, the corrosion can be detected by measuring the voltage between the metal and a reference electrode (RE), referred to as a half-cell electrode. The most well-known RE used in the industry is copper-copper sulfate (CSE). However, the common challenges facing underground RE signals is the high sensitivity to temperature and humidity swings. The voltage measurement is termed the corrosion potential voltage. The potential measurements provide a quick indication to evaluate corrosion between two different metals in an identified environment in neutral soil and seawater.

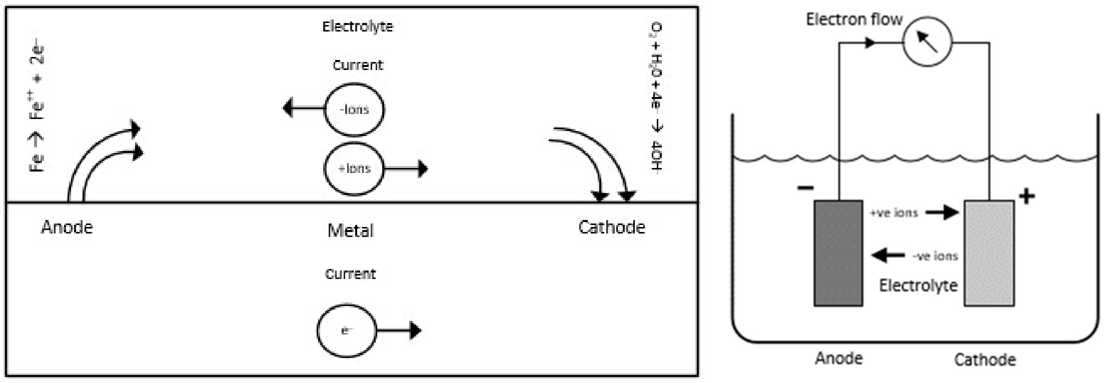


Figure 1. Scheme drawing of a basic corrosion cell (*left*), and Four necessary components of a corrosion process. (*right*)

For simplicity, the potential voltage for any metal or alloy, from figure 1 (*left*), to CSE is a direct measurement of the relative resistance to corrosion. However, if two dissimilar metals or alloys, from figure 1 (*Right*), are electrically connected in a given environment, the more positive member (noble member, i.e. lead) will act as a cathode and the more negative member (active member i.e. zinc) will act as the anode. In real applications, such as underground pipelines, the anodic and cathodic parts exist on pipelines. The electric current (electrons) flows from the anodic areas through the surrounding electrolyte to the cathodic areas on the pipeline. The anodic areas are where the corrosion is present. Using this information, the CP technique involves simply connecting a very active metal to the pipeline in order to reverse the current flow to make the pipeline act as a cathode and the active metal as the anode (termed the sacrificial anode). In this case, the corrosion rate does not stop, but rather the corrosion is displaced to the active sacrificial anode and concurrently protects the pipeline (Peabody, 1967).

Table 1. Material galvanic series in neutral soils (Peabody, 1967).

Material	Potential Volts (CSE)	Material	Potential Volts (CSE)
Platinum	- 0.10	Aluminum Alloy	- 1.05
Lead	- 0.50	Zinc	- 1.10

4. IoT concept

The Internet of Things (IoT) is a concept that has been emerging recently to a significant degree. It enables things (objects) to communicate, corporate, and interact with each other through wireless or wired connections. A significant amount of research has been carried out globally to make things more digitalized and virtual in order to overcome the various challenges associated with ensuring that environments are smart and more intelligent (Vermesan et al., 2013). In 2011, devices that were connected to the internet exceeded 7 billion (more than the Earth’s population at that time) and is expected to rise to 50 billion by 2020 (James, 2014). A study by General Electric (GE) in 2018 states that 64% of the utilities and power companies in North America rely on IoT applications (GE, 2017). IoT is a concept that is not based on one technology, but rather involves relating and connecting multi-things (multi-systems) such as traffic lights and implanting them with image recognition functionality (IRF) and sensors that are then integrated into analyst and decision support along with asset management. This completely integrated system forms the basic concept of IoT, which brings to the surface entirely new business opportunities and a more complex level of IT (Cooney, 2011). The concept of IoT is being heavily researched in other areas such as transportation, reverse logistics, logistics, and distribution where a significant amount of information has been integrated, communicated, and processed to make these systems more efficient and cost effective. As a result, the requirement for highly cognitive technologies and contextual intelligence will be essential in the near future for such new integrated systems. IoT existing architectures vary because they are vast and broad. However, such architecture typically comprises sensors, communications, computing technologies, and networks (Gigli & Koo, 2011). Further descriptions of IoT architecture's key layers and consistence are provided in Ref. (Zhang, Sun, & Cheng, 2012).

5. Cathodic protection integration model based on IoT

5.1. IoT architecture

In this study, the IoT architecture (figure 2) proposed was based on Wireless Sensor Networks (WSN) and energized by a power source. The sensors transfer data to an appointed center through Wi-Fi. The sensors are volt-ohm-milliamp (VOM) to measure the volts potential or Ampere current flow from the sacrificial anode to the underground pipe. The volts' fluctuation range reading should be within NACE standards (Fitzgerald Iii & Fnace, 2014). Sensors can also explore and detect broken wires, bad wire connections, or a broken weld. All these data will be transferred through Wi-Fi using IoT.

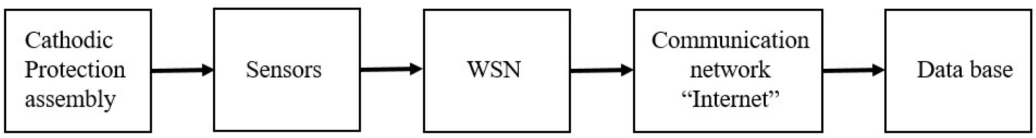


Figure 2. IoT Architecture

5.2. Cathodic protection components description

CP involves two main approaches, using either (1) galvanic anodes or (2) an impressed current. This paper focuses on the first method, due to its simplicity. From section 3, we discussed how a corrosion cell is set by contacting dissimilar metals. One metal of the corrosion cell forms the active side, which will corrode eventually (termed the sacrificial anode), while the other metal forms the noble side. From the aforementioned example, we will demonstrate an installation of CP with sacrificial anodes to buried metal pipelines in soil. This real application will be used throughout this paper, yet the notion of this paper is applicable for many other CPs with sacrificial anodes existing, such as underground tanks and lower standpipes of a fire hydrant. The buried pipeline is protected by connecting a very active metal (zinc or magnesium) to the pipeline with a wire (figure 3). The active metal will discharge electrons (current) to the buried pipeline through the wire in order to reverse the existing normal corrosion cell on pipelines. The sacrificial anode in such installation protects significant distances of a pipeline. Therefore, for long distance pipelines, sacrificial anodes are placed at specific points along a pipeline termed hot spots. The sensors might be distanced apart based on several factors such as the national regulations, NACE standards, corrosion location historylog, and high-risk corrosion spots. Once the sacrificial anode depletes due to corrosion, the corrosion begins on the pipeline. Therefore, frequent inspection and evaluation of the sacrificial anodes is crucial.

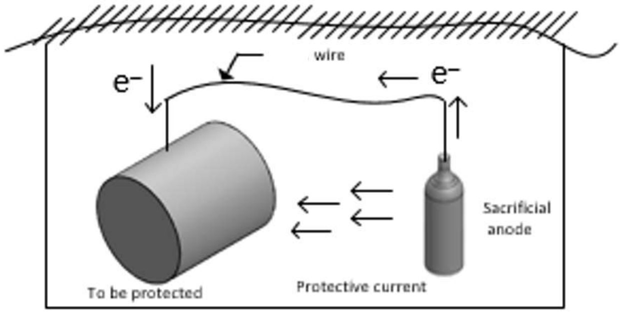


Figure 3. CP of buried pipelines in soil.

6. Proposed system setup

6.1. Block diagram

Different methods may be used in the proposed model process, including machine-to-machine technology (M2M), IoT, and cloud computing. M2M allows two technologies of the same type to relate and communicate. M2M communication can be done through wires or wirelessly. It is also considered to be an inherent part of IoT that can provide various gains to business and industries for control and monitoring purposes. The core controller used in this monitoring system is an Intel Galileo Gen2 board. This board is used due to it providing several advantages: it can transmit data wirelessly in the IoT environment, it can be connected to many types of sensor networks and sensors types, it can endorse diverse features of algorithms for future developments and improvements, and it can monitor and control parameters remotely. The only parameter used in this system is the corrosion potential voltage that checks the corrosion rate between the metal and a reference electrode. The corrosion potential voltage shows whether or not the sacrificial anodes need to be replaced. The block diagram of the system has the following components: corrosion rate sensor, Intel Galileo Gen2 board, and Wi-Fi module (as demonstrated in figure 4). The block diagram of the proposed monitoring system shows the development of measuring the corrosion rate in a cathodic protection system. The sensor data is gained by the Intel Galileo Gen2 board, processed, and sent to the assigned department by the IoT concept.

Figure 4. Proposed block diagram.

6.2. Flow chart

The system receives the measured attributes (voltages) from different sensors along a buried pipeline. This is done by the Intel Galileo Gen2 board, the interface device, which can discover any sensors linked to it. It can also control sensor data acquisition, data processing, and data transmission. The readings are then sent to an authorized person through Wi-Fi for further analysis and decision-making. A flow chart of the proposed system is shown in figure 5 (*left*)

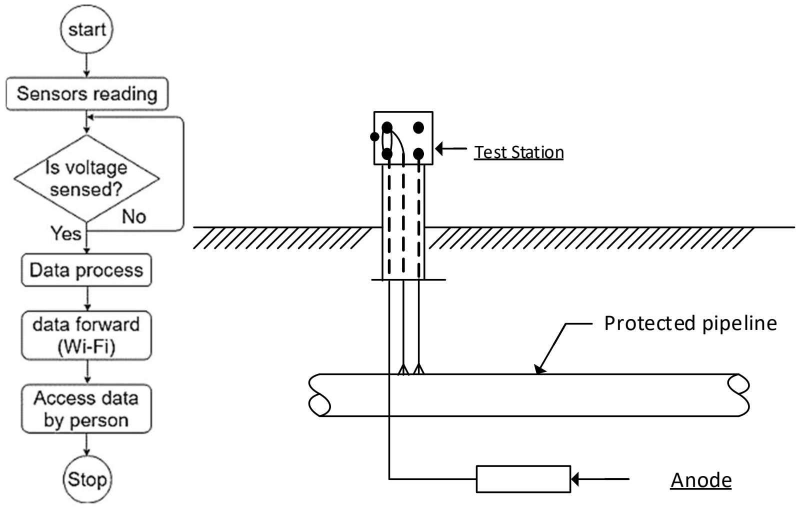


Figure 5. Flow chart of proposed system (*left*) and CP rate monitoring insulation setup (*Right*).

6.3. Cathodic protection monitoring technology and sensors

Many options are available for remote monitoring technologies such as SCADA-based (Supervisory Control and Data Acquisition), Satellite-based, and Smart phone-based systems. All these monitoring technologies are suitable and compatible for effectively monitoring cathodic protection systems. The latter technology is the most effective and economical method in this regard. The smart phone-based system is known for its low cost and ease of installation, but carries with it a certain level of risk because its signal can be lost in locations with low reception. The system will display the voltage potential measurements through sensors that determine the corrosion potential parameters. Many types of corrosion sensor can be used for this project such as Corrosion Chip (CC) probe, Multielectrode Array Sensor (MAS) Probe, corrosion Resistance probe, and the voltage measurements sensor. The voltage measurements sensor (units: mV) is preferred to have the lowest practical range for accuracy reasons. NACE SP0169 states that the accurate mV measurements of a CP system applies in the upper two-thirds range of the voltage sensor (NACE, 2013). Based on the reverse potential calculations of the designed CP system, the voltage must be in the range of -0.805 to -0.895 in order to ensure that the system is protecting the underground pipelines. Any voltage potential readings outside of the stated range indicate sub-optimal CP performance.

7. Experimental Tests and Results

The sensed data (parameters) comprises the corrosion potential voltage, in milli-volts, between the metal pipeline and selected sacrificial anodes. The pipe is made of steel and the selected sacrificial anode is magnesium. The uncoated steel pipe is buried in moist soil to enhance the corrosion medium. The corrosion sensor is connected to the steel pipe and to the sacrificial anode to measures the corrosion potential voltage in moist soil medium, as shown in the schematic drawing in Figure 5 (Right). The wires and communication circuits were well coated and insulated to avoid any environmental impact such as water or heat. Relatively simple hardware was implemented using C++ function coding in an Arduino IDE. The results of acquired data of the corrosion potential voltage (CPV) are shown in figure 6 (left). The CPV readings in the first days (45 days) were within the range of an appropriate level of CP performance. However, and in order to test a different scenario in fewer days of the CP system, sulfides and acids were added to the moist soil. This alternation in the experiment will lead to a change in the CPV that result in more degradation and depletion in the sacrificial anode rate and thus interrupts CP performance. This change is observed clearly in the CPV curve readings after day 45 in figure 6 (left & right) on the date Sep 14, 2018. The corrosion rate sensor forwards data to the Intel Galileo Gen2 board in order to be processed. The processed data are then sent by Wi-Fi to an authorized person to further data analysis. This reported data is shown in figure 6 (right).

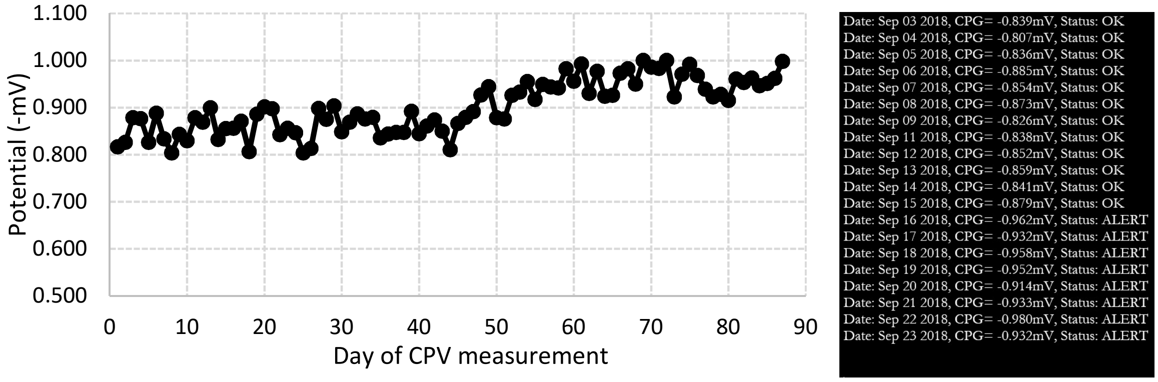


Figure 6. Daily CP system voltage potential readings (*left*) and CPV readings on mobile screen connected to the Internet (*Right*).

8. Conclusion

This study proposes a design of a smart sensor interface for monitoring Cathodic Protection (CP) in an Internet of Things (IoT) environment. A corrosion sensor is used to monitor voltage potential from a sacrificial anode to an underground pipe or pipelines. The interface device is the crucial part of the proposed system that can control sensor data acquisition, data processing, and data transmission. The Intel Galileo Gen2 board provides several advantages: it can transmit data wirelessly in the IoT environment, it can be connected to many types of sensor networks and sensors types, it can endorse distinct features of algorithms for future developments and improvements, and finally it can monitor and control parameters remotely. The recent advancement and improvements in IoT technologies have led to a revolution in wireless communication in different applications, from security monitoring to engineering management. The past design of monitoring CP was via wired networking (for data transfer) or by onsite inspection. Although such inspection provides a highly reliable monitoring method, it is expensive and inconvenient. Wireless technologies in IoT environments are used to reduce cost and maintain the same level of reliability. The results and performance of the proposed design have been tested and reported. The Intel Galileo gen2 board is simple and timesaving to use, in addition to it already being equipped with a system that can easily be connected to the internet. The proposed system is designed to monitor CPV and reveal the performance of a CP system in any industrial application. The design of the monitoring application on a smart phone is very simple and significantly less costly than other monitoring technologies. Considering the impressive experience and results of this proposed system, we may apply the same approach, in future, to other isolated systems in refineries to convert them into smart-refineries.

9. References

Chi, Q., Yan, H., Zhang, C., Pang, Z., & Da Xu, L. (2014). A reconfigurable smart sensor interface for industrial WSN in IoT environment. IEEE transactions on industrial informatics, 10(2), 1417-1425.

Co., G. E. (2017). GE Digital Industrial Evolution Index. Retrieved from GE 2017 Report.

Cooney, M. (2012). Gartner: The top 10 strategic technology trends for 2012. Network World.

Fadell, A., Rogers, M., Satterthwaite, E., Smith, I., Warren, D., Palmer, J., . . . Fiennes, H. (2017). Smart home system using motion sensing for occupancy determinations and display screen activation determinations. In: Google Patents.

Fitzgerald Iii, J. H., & Fnace, M. P. (2014). Safety First.

Gigli, M., & Koo, S. G. M. (2011). Internet of Things: Services and Applications Categorization. Adv. Internet of Things, 1(2), 27-31.

James, R. (2014). The internet of things: a study in hype, reality, disruption, and Growth. Raymond James US Research, Technology & Communications, Industry Report.

Koch, G. H., Brongers, M. P. H., Thompson, N. G., Virmani, Y. P., & Payer, J. H. (2002). Corrosion cost and preventive strategies in the United States. Retrieved from

Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. Business Horizons, 58(4), 431-440.

Minoli, D., Sohraby, K., & Occhiogrosso, B. (2017). Iot considerations, requirements, and architectures for smart buildings—energy optimization and next-generation building management systems. IEEE Internet of Things Journal, 4(1), 269-283.

NACE, S. (2013). Control of External Corrosion on Underground or Submerged Piping Systems.

Peabody, A. W. (1967). Control of pipeline corrosion: National Association of Corrosion Engineers.

Porkodi, R., & Bhuvaneswari, V. (2014). The internet of things (IOT) applications and communication enabling technology standards: An overview.

Ratnayake, R. M. C., & Kusumawardhani, M. (2013). Reliability analysis of condition monitoring data on aging plants: A case study from topside static mechanical systems.

Salunke, P., & Kate, J. (2017). Advanced smart sensor interface in internet of things for water quality monitoring. Paper presented at the International Conference on Data Management, Analytics and Innovation, India.

Vermesan, O., Friess, P., Guillemin, P., Sundmaeker, H., Eisenhauer, M., Moessner, K., . . . Cousin, P. (2013). Internet of things strategic research and innovation agenda. RIVER PUBLISHERS SERIES IN COMMUNICATIONS, 7.

Zhang, M., Sun, F., & Cheng, X. (2012). Architecture of internet of things and its key technology integration based-on RFID.

Big Data, Ethics, and Social Impact Theory – A Conceptual Framework

Gwen White[1]
Thilini Ariyachandra[1]
David White[1]
[1]*Management and Information Systems and Psychology*
Xavier University
whiteg@xavier.edu

Abstract

Decisions made using big data, impact ethical issues like privacy, security, ownership, and decision making. In addition, those same decisions can have a positive or negative social impact. This paper proposes a framework that explains how decisions made using big data impact ethics and social impact theory. A broad literature review explored how big data and ethics can have a social impact on society. The proposed framework of big data, ethics and social impact is illustrated through three examples. Insurance companies manipulate big data to impact sales. The Center for Disease Control examines big data to determine the location of the next outbreak. Companies analyze big data in predictive analytics to increase marketing or determine a new trend. It was found that these uses of big data directly affect ethics which has a positive or negative social impact. Simple decisions can change the outcome for one or millions.

1. Introduction

Big data is a disruptive technology where it changes an existing industry or starts something completely new (Wessell, 2016). Big data is found in many industries like finance, education, government, and even retail. (Disruptive Technology Reconsidered: A Critique and Research Agenda, 2004). There are many types of big data which include mobile, pictures, social media, video, machine generated and audio. These new types of media will make up 85% of data sources for big data consistently increasing the size of the big data cache (Bearden, 2014). Within the past ten years, ethical problems associated with big data have increased. Big data affects the ethical principles of privacy, security, ownership, and decision making. Many organizations use the results from big data analytics to make major decisions. Those decisions directly influence ethics and in turn, play a role in the social impact of its outcomes. This paper proposes a framework that explains how big data, ethics and social impact theory are interrelated.

2. Big data and ethics

There is a rapidly growing volume of literature that combines big data and ethics. Between 2001 to 2016 big data usage has increased along with the instances of actual and perceived ethical violations. Ethical themes evolved including privacy, security, ownership, and decision making. According to Zwitter (2014), industry is moving towards "changes in how ethics has to be perceived away from individual decisions with specific and knowledgeable outcomes towards actions by many unaware that they may have taken actions with unintended consequences for anyone". There are four themes that have emerged in relation to ethics and big data: privacy, security, ownership, and evidence-based decision making.

Privacy is the non-disclosure of personal information to the public (Davis, 2016). Big data can contain private information that might be exposed unintentionally or intentionally. States normally develop their own privacy requirements which echo federal laws (Thorpe & Gray, 2015). Normally data should be de-identified to protect it. Since so many people are connected through a variety of data networks, the ability to generate and share data increases daily (Tene & Polonetsky, 2013). There are situations where consent is not possible. “Protecting privacy will become harder as information is multiplied and shared ever more widely among multiple parties around the world. As more information regarding individuals’ health, financials, location, electricity use, and online activity percolates, concerns arise regarding profiling, tracking, discrimination, exclusion, government surveillance and loss of control” (Tene & Polonetsky, 2013, pg. 2).

Security is designed to protect data from others who do not have access permission. As more data is produced, there are more opportunities for data breaches. The various software servers and locations are vulnerable to hacking because many processors are located outside the jurisdiction of the company that requested the information (Thorpe & Gray, 2015). The overwhelming amount of big data makes it hard to protect and the protections are not always sufficient. The lack of protection for data is a violation of ethics because it affects more than just security. The public wants to feel confident their data is protected and secure.

Data ownership is a feeling of control but most people do not own their data, they think they do. Big data is harvested from a variety of locations and organizations that receive the data assume it has been responsibly acquired. Just because the form has an opt-in or opt-out checkbox does not mean it will be respected. The individual does not own the data. It is just a mere pit stop before hitting the superhighway of data transmission known as big data.

Organizations use data to make decisions about populations regarding what they should receive or not receive. Many times, algorithms are used to determine needs but at the same time these decisions can affect a population because personalization is removed. Predictive analysis, making decisions based on a set of parameters, is common but are these decisions appropriate? “The wealthy and well-educated will get the fast track; the poor and underprivileged will have the deck stacked against them even more so than before” (Tene & Polonetsky, 2013, p. 254). Decisions based solely on numbers can be problematic and lead to discriminatory practices by ignoring the personal aspects of the data.

Ethics and big data are intertwined and in combination they have a social impact. New ethics rules are starting to emerge that will help organizations and individuals with responsible use of big data, however, there are social consequences that are evidenced in social impact theory.

3. Big data and social impact

Big Data can be used to help address various social problems related to societal issues such as hunger, disease, poverty, and social inequity. Social problems are often what are called “wicked” problems (Ritchey 2011). Not only are they messier, but they are also more dynamic and complex than technical or business problems as the number of stakeholders involved is high. While nonprofits and government agencies may take part in resolving social problems, collaboration and data sharing is often the exception rather than the norm. During social crises, the data collected may not be effectively integrated to serve all effected constituents immediately. The majority of big data used in the context of social crises and disasters is called big crisis data (Castillo 2016). The data driven decision making resulting from the analysis of big data can result in far reaching societal impacts.

Social impact is a nebulous concept that over the years has been conceptualized in multiple ways

(Epstein et al 2014). A popular definition of social impact presented by Clark et al. (2004) is a portion of the total outcome that happened as a result of a decision or activity of an organization, above and beyond what would have happened anyway. Positive social impacts resulting from big data insights can range from 'better health, greater financial inclusion, and a population that is more engaged with and better supported by its government.' (LaValle et al 2011).

While the vast quantity of data gathered during a crisis can be very useful if subjected to instant analysis and recommendations to enable effective responses, it can also be paralyzing as most humanitarian organizations lack the infrastructure and the capability to conduct big data analysis. In addition to the large quantity of data collected, big data analyzed to create a positive social impact suffers from authenticity issues. Misinformation and disinformation generated on social disasters can influence analysis and recommendations from big data. Digital humanitarianism suffers from the collection of disingenuous data but also reveals the power of data by giving organizations the ability to create solutions for social problems that were previously not viable.

In addition, large historic data sets could lead to false confidence in predicted outcomes that may result in significant errors (Ganore 2012). Individual agendas of corporations or governments could lead to manipulated findings that in turn suggest the use of skewed recommendations for the selfish benefits of a few. Social impact theory provides a lens to consider how big data insights could lead to positive or negative social impacts and the ethics behind it.

4. Social impact theory

Social impact theory was created by Bibb Latané in 1981 at Ohio State University. This theory was defined as "any of the great variety of changes in psychological states and subjective feelings, motives and emotions, cognitions and beliefs, values and behaviors, that occur in an individual, human or animal, because of the real, implied, or imagined presence or actions of other individuals" (Latané, 1981, p. 343). The theory uses mathematic equations to determine the various levels of social impact based on the situation.

The social influence that is predicted will have a proportional multiplicative influence and the number of people involved in the social influence will have an inverse proportional influence on the number of people influenced (Latané,1981). People's actions affect others in social situations and the impact of their actions can be measured visually along with measures that include three laws: social forces, psychosocial law, and multiplication/division of impact (Table 1).

Table 1. Social Impact Theory

Social Force	• Strength of the message • Immediacy • Number of people/target influence
Psychosocial Law	• First source influence most important
Multiplication/Division of Impact	• Smaller impact less dispersion • Larger impact more dispersion (Latané, 1981)

Social force is a pressure that is put on people to change their behavior. The concept of social force is divided into three sections: strength of the message, immediacy, and numbers of people exerting force on a group (social force = f(S*I*N)) (Latané, 1981). The strength of the message is based on the level of power and influence perceived by the target. Immediacy concerns how recent the event occurred and if there were other factors involved. Number of people includes the total number of influencers on the target. Psychosocial law explains that the first source of influence is the most potent. Multiplications/divisions of impact emphasize that the force directed at one person will have a large impact and if there are two or more people the impact splits proportionally (two people = impact split in half). This theory along with big data and ethics is applied to a conceptual framework that describes social impact based on decisions that use big data.

5. Big data, ethics, and social impact

Decisions using big data affect ethics and social situations. It is important to consider how and who is affected by decisions. Does the decision affect the public, is it detrimental, is it beneficial and does it create group dynamics? We propose that the use of big data has a direct effect on social impacts and ethics. The big data, social impact, and ethics conceptual framework are explained and represented in Figure 1. An organization gathers big data from sources. The big data is then delivered to a group or individual which processes the data by absorbing the data. That big data has an influence on the group or individual which can be positive or negative. Once the influence is determined, social impact is assessed. The level of social force, psychosocial law and multiplication/division of impact is assessed to determine the social impact. The level of social impact is directly compared to ethics to determine if the social impact was ethical and its end effect on society.

Depending on how big data is used, it could create fear or euphoria. The fear or euphoria can be used to initiate obedience and a need for immediate action. The action can either grow or reduce the number of people involved and be responsible for behavioral changes in the affected groups. The impact of the behavioral changes applied directly affects the social impact and ethics of the situation. The big data, social impact, and ethics framework conceptual model will be applied to show how these concepts are related.

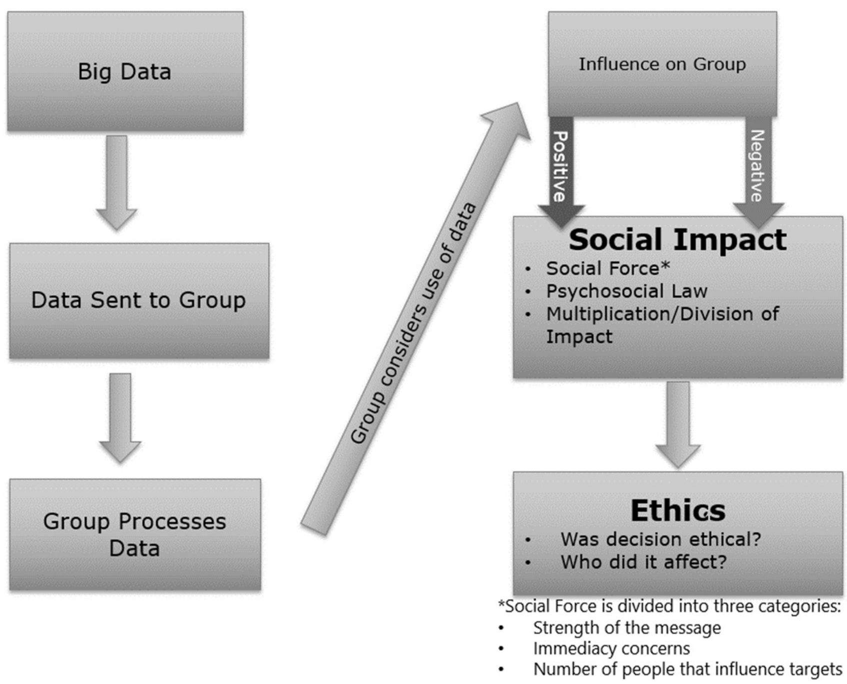


Figure 1. Big data, social impact, and ethics conceptual framework

6. Application of big data, social impact, and ethics

Big data is powerful and unpredictable. Important trends and new connections can be discovered when evaluating big data. In combination with predictive analysis, big data is an important tool for discovering new trends, new products and preventing disease. However, one must be careful when using big data because of ethical implications and their social impact. It is important to use big data responsibly. Three applications of big data, ethics and social impact are summarized next.

Insurance companies use tornado data to predict the likelihood of recurring storms in the same location. The data generated from the National Oceanic and Atmosphere Administration (NOAA) tracks various weather incidents throughout the United States. Moore, Oklahoma is a hotspot for tornadic activity. Moore has been impacted by ten tornadoes from 1998 to 2015. Since the city is in tornado alley, it is vulnerable to repeat incidents. Insurance companies determined that since Moore was a likely target, it was important to explain the need for homeowner's insurance coverage for protection from a potential tornado. The direct impact of the big data helped insurance companies deliver a message of safety, importance, and vigilance (Palmer, 2013). The insurance companies decreased cases of fraud, enhanced security and encouraged the community to help rebuild, make donations, and provide emotional support to the victims of the latest tornado. Their consumers were prepared for the next tornado. Social impact theory was applied through social forces (the significant influence of the insurance company), immediacy (the past tornadoes occurred in Moore) and the multiplication/division of influence (many people were informed of the insurance companies warnings). Their direct influence on groups of people made the community come together and help each other even more than before (Palmer, 2013). The ethically social impact was positive because the perspective of insurance companies, in this case, was changed from negative to positive.

The Centers for Disease Control (CDC) use big data to predict the spread of the flu. The review of data determines where the next outbreak might occur. These predictions can help doctors prepare for the flu outbreak and in addition encourage individuals to get flu shots for preventative measures. The use of big data, in this case, saves lives. Social impact theory was applied through social forces (the CDC had a significant influence on the community), immediacy (the importance of the flu epidemic) and the multiplication/division of influence (impact of many people who could be affected by the flu). The CDC has a strong influence over the public, therefore any messages delivered by the organization are impactful and important. Decreasing the spread of disease is ethical and necessary (Silva, 2006). It is vitally important to maintain the public's trust, especially when lives are at stake.

Big data is used to recommend products to consumers through predictive analysis. However, companies should be careful when recommending a product due to sensitivity and ethical problems. An e-cigarette company can have a strong influence on the consumer (Zhu, et al., 2014). The e-cigarette company can make recommendations that are both positive and negative, potentially creating an ethical dilemma. For example, big data has been used to determine if smokers might be interested in e-cigarettes. Unfortunately, smoking has caused many deaths (Center for Disease Control, 2016). However, the e-cigarette companies see e-cigarettes as a safer alternative for their customers. Social impact theory was applied through social forces (the e-cigarette companies have a strong influence on the existing smoking customer base), immediacy (a brand-new product was introduced as an alternative to tobacco cigarettes) and multiplication/division of influence (the information was shared with a large customer base who influenced each other to purchase a new product). Due to the strong influence of the company, people made the change and in some cases, the consumers convinced others to change based on the message from the company. In this case, the promotion of negative behavior in a positive light is questionable and depending on the audience, violates ethical principles.

7. Conclusion

Big data usage has increased and we should watch for ethics violations and negative social impacts. This article examined how big data, ethics and social impact are interrelated using a conceptual model. It is important that organizations be respective when using big data to make decisions because it directly impacts individuals and society. The framework was supported through three different examples that indicated big data can be used to manipulate ethics and social impact. Organizations must make the right decisions, honor established ethical principles, and make positive social impacts. A simple decision is enough to change the outcome for thousands, even millions.

8. Future Work

Further examination of the Big Data, Ethics, and Social Impact Framework is needed to determine the exact impact. The next step is to evaluate the validity of the big data, ethics, and social impact framework through semi-structured interviews with data analytics managers and consumers along with empirical survey research which will statistically validate the conceptual model.

9. References

Bearden, R. (2014). Hadoop Summit. CEO Keynote Conference Speech. San Jose, CA: Hadoop Summit.

Center for Disease Control. (2016). Smoking and tobacco use. Centers for Disease Control.

Davis, J. (2016, January 4). Big data predictions for 2016. Retrieved May 5, 2016, from Information Week: http://www.informationweek.com/big-data/big-data-analytics/big-data-predictions-for-2016/d/d-id/1323671

Disruptive technology rconsidered: A critique and research agenda. (2004). Journal of Product Innovation, 21, 246-258.

Ganore, P. (2012, October 1). Need to know what Big Data is? Retrieved April 2, 2017, from ESDS Enabling Futurability: https://www.esds.co.in/blog/need-to-know-what-big-data-is/#sthash.PQpIh0YO.dpbs

Johnson, D. (1985). Computer ethics. Englewood Cliffs, NJ: Prentice-Hall.

(2014). Joint Statistical Meetings.

Latané, B. (1981). The psychology of social impact. American Psychologist, 36(4), 343-356.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data analytics and the pathfrom insights to value. MIT Sloan Management Review, 52(2), 1-15.

Needham, J. (2013). Disruptive possibilities: How Big Data changes everything. Boston, MA: O'Reilly Media.

Olumoye, M. Y. (2013). Ethics and social impact of information systems in ethics and social impact of information systems. International Journal of Science and Research (IJSR), 2(11), 154-158.

Palmer, J. (2013, May 22). Oklahoma tornadoes: Insurance companies set up in Moore, in it for the long haul. Retrieved March 31, 2017, from The Oklahoman: http://newsok.com/article/3828511

Richards, N., & King, J. (2014, May 19). Big Data ethics. Wake Forest Law Review, 49, 393-432.

Silva, M. L. (2006, November 6). What would you do? Ethics and infection control. Online Journal of Issues in Nursing, 12(1), 1-3.

Tene, O., & Polonetsky, J. (2013, April). Big Data for all: Privacy and user control in the age of analytics. Northwesten Journal of Technology and Intellectual Propery, 11(5).

Thorpe, J. H., & Gray, E. A. (2015, March-April). Big Data and public health: Navigating privacy laws to maximize potential. Law and the Public Health, 130, 171-175.

Wessell, M. (2016, January 27). How Big Data is changing disruptive innovation. Retrieved May 6, 2016, from Harvard Business Review: https://hbr.org/2016/01/how-big-data-is-changing-disruptive-innovation

Zakir, J., Seymour, T., & Berg, K. (2015). Big Data analytics. Issues in Information Systems, 16(II), 81-90.

Zhu, S.-H., Sun, J. Y., Bonnevie, E., Cummins, S. E., Gamst, A., Yin, L., & Lee, M. (2014). Four hundred and sixty brands of e-cigarettes and counting: Implications for product regulation. Tobacco Control, 23(3), 3-9.

Zwitter, A. (2014). Big data ethics. Big Data and Society(July - December), 1-6.

Performance Enhancing Drug Usage: The Influence of Adverse Health Effects and Public Embarrassment

B. Andrew Cudmore[1]
Sherry Jensen[1]
[1]*Florida Institute of Technology*
acudmore@fit.edu; sjensen@fit.edu

Abstract

This research examines the impact of adverse health effects and public embarrassment as deterrents to the use of performance enhancing drugs (PEDs). Deterrence theory suggests that potential PED users execute a cost-benefit analysis before engaging in illicit drug use; any increase in perceived costs reduces the likelihood of drug use. In accordance with the deterrence theory, this study finds that social costs (public embarrassment) have a negative impact on attitudes toward PEDs. However, potential health costs, even extreme ones, do not deter amateur athletes from considering PEDs. Rationale is offered for why fear of social disapproval has a larger impact than adverse health outcomes on attitudes of potential PED users. Results provide guidance for the development of marketing communications designed to deter amateur athletes from considering PEDs.

1. Introduction

Valued at $60.5 billion in 2014, the North American sports market, as captured by gate revenues, media rights, sponsorship and merchandising, is projected to reach $73.5 billion in 2019 (Heitner, 2015). Further, youth sports in the United States are a $17 billion market (Jimenez, 2018). Sport in modern society has evolved into a multi-billion dollar industry such that athletes are ever-present in the media and sports celebrities become powerful role models. This commercialization of athletic competition has created an incentive structure that increases the payoff from the use of performance enhancing drugs (PEDs). However, the use of PEDs is both illegal and unethical.

Drug use by athletes is defined as illegal by the World Anti-Doping Association if the drug enhances performance, risks the health of the athlete, or violates the "spirit of sport" (World Anti-Doping Agency [WADA], 2015). Also referred to as PES or performance enhancing substances (Momaya, Fwal and Estes, 2015), PED use dates back as far as the third Olympiad and the International Amateur Athletic Federation imposed the first official ban on "stimulating substances" in 1928 (Savulescu, Foddy, & Clayton, 2004). As early as the 1960s, the problem of PED consumption by professional athletes has been acknowledged as problematic and in 1967, the International Olympic Committee formed a Medical Commission as a reaction to the increasing utilization of PEDs (CNN, 2019). The National Football League (NFL) began to test players for steroids in 1987, and in the following year, the Anti-Drug Abuse Act was passed by the United States Congress (CNN, 2019). A decade later the World Anti-Doping Agency was created in 1999, followed by the United States Anti-Doping Agency in 2000 (CNN, 2019). The 21st Century has seen federal investigations and congressional testimony on the topic of PED use, athletes stripped of titles and banned from competition due to the use of prohibited substances, and underground steroid labs closed in drug busts and distributors imprisoned (CNN, 2019). The problem of PEDs is pervasive.

Given the negative social aspects of this problem, it is proving difficult to assess its magnitude. For example, although official aggregate studies from a variety of sources (including the World Anti-Doping Agency, Australian Sport and Anti-Doping Agency, and the United States Anti-Doping Agency) estimate the proportion of athletes violating drug use policies to be 2% or below, incidence rates rise sharply when anecdotal and unofficial estimates are considered (Quirk, 2009). Today there is the perception that almost all elite athletes engage in PED use (Chan, Dimmock, Donnovan, Hagger, Hardcastle, & Lentillon-Kaestner, 2014; Kisaalita & Robinson, 2014; Savulesc et al., 2004; Ungerleider, 1992). Given the adverse health effects of PED usage and the stance that PED usage violates social ideals, concern grows as PED use permeates the sport culture and successful athletes influence amateur athletes.

Since the development of the anti-doping position in the 1960s, efforts to police PED use have primarily utilized a detection-based deterrence strategy (Chan et al., 2014). Weaknesses of this detection strategy have led to the exploration of prevention-based deterrence strategies, which focus on social forces (Mazanov, 2009). Arguably, amateur athletes provide a great target audience for prevention-based communications, especially if they can be reached before they start using PEDs. Amateur athletes see the potential for opportunity (money and fame), yet they may not see the cost (health and social) or do not feel it is personally relevant. Given that PEDs increase the health risks of an athlete (Pope, Wood, Rogol, Nyberg, Bowers, & Bhasin, 2014; WADA, 2009) and are also legal and ethical violations, the use of PEDs should be discouraged (WADA, 2015). One way to deter the use of PEDs is to educate athletes on the costs involved. However, learning is more apt to happen if a message is attended, salient, understood, and is personally relevant (Howe & Krosnick, 2017). Therefore, costs must also be perceived as true costs by these athletes. This study tests two general categories of perceived costs (health and social) to determine their effect on attitudes of amateur athletes toward PEDs. From this, it can be determined what costs should be emphasized in marketing communication designed to deter the use of PEDs by amateur athletes.

Section 2 presents the theoretical foundation and motivates the hypotheses, followed by the study design (section 3), results (section 4), and conclusions, contributions, limitations, and future research (section 5).

2. Theoretical foundation

2.1. Prevailing theories

Designing marketing communications to influence amateur athletes presents a challenge. First, information must be present and attended before it can be utilized to make a decision. In order to be attended, the message must be salient. Salience increases with involvement and personal relevance (Grover & Miller, 2014). This study endeavors to assess whether certain health and social costs are attended (personally relevant), and then influential in determining amateur athletes' attitudes toward PEDs. From this, we can determine what perceived costs should be showcased in marketing communication targeted at amateur athletes. If the message is deterrence of PEDs, then deterrence theory, which is rooted in the rational-choice perspective of human behavior (e.g. the theory of planned behavior), asserts that rational actors weigh the costs and benefits associated with an action before executing a decision (Loughran, Paternoster, Chaflin, & Wilson, 2016; Strelan and Boeckmann, 2006).

Second, information used in decision making can come from many sources. The TRA suggests that reference groups and their social norms are important (Ajzen, 1991). These groups may include peers (fellow athletes), coaches, family (pushing young athletes to succeed), or aspirational groups (athletic heroes). In turn, amateur athletes recognize such opportunity demands certain costs to be competitive. The question remains as to what costs they are willing to pay and why. Third, information is not always processed objectively, therefore the focus is on perceived costs and benefits.

Fourth, how information is weighted also depends on its assigned probabilities of occurrence. For example, short term benefits may be seen to outweigh long term costs, especially if they are not certain costs (Murphy & Dennhardt, 2016). This may be enough to discount a cost in the mind of an amateur athlete.

2.2. The Importance of perceived health costs

There is a wide range of health consequences that can result from PEDs. In fact, physicians have been called to better educate themselves on PEDs and their possible outcomes and in turn educate athletes (Momaya, Fwal and Estes, 2015). Strelan and Boeckmann (2006) cite research that suggests the propensity to engage in PED use is reduced by educating athletes on the associated health risks. In turn, it is expected that perceived health costs would be attended if they are salient. However, it is important to note that many athletes, and particularly amateur athletes, may not know all the forms of PEDs (anabolic drugs, stimulants, ergogenic aids, human biomolecules, adaptogens, nootropics, pain killers, sedatives, anxiolytics, blood doping, or gene doping), let alone the health consequences of taking them. This clearly motivates the need to educate athletes.

Research has revealed a host of potential health complications, both physical and mental, and in some cases the risk of mortality; however, media attention emphasizes the advantage gain from PED use, not the health risks (Pope et al., 2014). So if an athlete is educated in the variety of forms of PEDs, and can, in turn, recognize them, it is expected that PEDs would represent a perceived health cost. However, extant research suggests that the influence of perceived health costs in the decision to use PEDs is mixed (Bloodworth and McNamee, 2010; Kirby, Moran, & Guerin, 2011; Quirk, 2009; Wiefferink, Detmar, Coumans, Vogels, & Paulussen, 2008). For example, if athletes view the outcome of PED use as success in competition and not the execution of a negative health behavior, PED use prevails as a natural tool to meet athletic goals (Quirk, 2009). This suggests that some athletes may only focus on the benefits, either ignoring the costs or discounting them. Such weighting could be related to the chances of the cost occurring versus the chances of the benefit (Andrews, 1998; Bloodworth and McNamee, 2010), or it could be based on perceived norms of reference groups that decide to participate or not in PEDs (Wiefferink, Detmar, Coumans, Vogels, & Paulussen, 2008). However, when the prize is big enough the costs may not be given much weight. Interviews of elite athletes across a variety of sports by Kirby et al. (2011) revealed that health consequences were not especially influential in their decisions to start using PEDs; possible health side effects were rarely cited as a concern. Interviewees in Bloodworth and McNamee's study (2010) openly acknowledged that elite competition has health consequences and that they prioritize winning over health. Some confident athletes even suggested that their body management skills reduced their risk. Nonetheless, it is expected that such health-conscious individuals as athletes would be salient of the risks and therefore, the cost to their health such that, based on the theory of reasoned action and perceived deterrence theory, the following hypothesis emerges:

> H1: The higher the risk of negative health effects, the more negative the athlete's attitude toward towards performance enhancing drugs will be.

2.3. The importance of perceived social costs

Amateur athletes are victims of a host of social forces (Connor, 2009). These forces (reference groups) can either promote or deter behavior. First, our pervading philosophy of capitalism is profit focused (fortune) and exalts winning (fame). Second, athletes encounter an incentive structure (fame and fortune) where success outweighs possible sanctions. Therefore, these social forces work to indirectly promote clandestine use of PEDs (given it remains hidden from public scrutiny) and only the outcome of athletic success (fame and fortune) is promoted. The focus here is on the social benefit. However, PEDs represent a social cost if they are revealed, as they are illegal and unethical.

A number of perceptual-deterrence studies, examining unlawful acts of varying degrees of severity, have shown that moral beliefs and fear of social disapproval are the most likely deterrents to illegal activity, not legal sanction, and the more the individual anticipates that disapproval will come from an important person, the less likely the individual will engage in the behavior (Strelan & Boeckmann, 2006). Research indicates that informal sanctions, including loss of friendships, employments, status, and trust in relationships, are a much stronger deterrent than formal sanctions (Overbye, Elbe, Knudsen, & Pfister, 2015; Kelly, Fukushima, Spivak, & Payne, 2009). Young athletes interviewed by Bloodworth and McNamee (2010) revealed that social and moral constructs were especially strong deterrents to the use of PEDs, with the impact of social sanctions and anticipated shame from PED use exposure as a theme present in focus group discussions. Interviews revealed to Kirby et al. (2011) that thoughts, emotions, and feelings were more impactful on athletes' decisions than stressing negative health consequences. Jalleh, Donovan, and Jobling's (2014) survey of 1,237 elite Australian athletes revealed that both morality and reference group opinion, along with the perceived legitimacy of drug testing and the appeals process, had a significant relationship with attitude towards the use of the PEDs. In light of their findings, it is suggested that anti-doping education would be most effective if it included discussion of moral decision-making behavior and the embarrassment if caught using PEDs. The TRA suggests that the greater the extent to which PED use is viewed as socially unacceptable (unsupported), the greater the motivation for athletes to comply with the social norm of non-use leading to the following hypothesis:

> H2: The greater the level of unsupportive media coverage (versus lower and supportive media coverage) revealing an athlete's PED use, the more negative the attitude an athlete will have towards performance enhancing drugs.

2.4. The importance of perceived health and perceived social costs

As discussed in Ntoumanis, Ng, Barkoukis, and Backbone's (2014) meta-analysis of personal and psychosocial predictors of doping, the use of PEDs is a complex decision and impacted by a number of both situational and personal variables. Given that both health costs and social forces are individually expected to decrease the propensity to engage in PED use, it is expected that when both health and social forces are at high levels, the deterrent to PED use will be greatest. For the latter, how the media reacts (level of coverage and whether supportive or unsupportive) will also influence perceptions. However, given prevention-based deterrence strategies focus on social forces (Mazanov, 2009), as supported by the TRA, we expect that social costs (as facilitated by media exposure) will outweigh health costs when calculating the cost/benefit of PED usage leading to the following hypothesis:

> H3: Under unsupportive, higher media coverage scenarios (versus lower, supportive media coverage), major health effects result in a more negative change in the athlete's attitude toward performance enhancing drugs, over minor health effects, compared to those found under low media coverage scenarios.

3. Methodology

3.1. Design and sampling

A 2 (minor/major health effects) X 2 (supportive and minor media coverage/unsupportive and major media coverage) ANOVA between subject balanced design utilized a stratified quasi-random method where only amateur athletes where surveyed. Specifically, every third amateur athlete leaving a campus athletic center was surveyed. The sample included 60% males and 40% female; 78% (18-25 age); 11% (26-35); 5% (36-45); 6% (over 45); and 72% American and an even distribution of other ethnicities 28%.

3.2. Procedure

After signing a consent form, participants were given one of four short scenarios to read and then rated their attitudes toward the use of PEDs. The four scenarios differed by the level of adverse health effects and media coverage as a measure of the potential for public exposure and embarrassment. All scenarios began with the same pictures of a young athlete in a baseball uniform followed by a "+" sign, then a picture of pills in a see-through bottle, followed by an "=" sign, and then a picture of a well-muscled adult version of the same, now famous athlete. These collective images were meant to imply the ability of PEDS to enhance an athlete's physique and career.

3.3. Dependent measure

The dependent measure was the average of a standard 3-item, 7-point semantic differential scale measuring attitude:

My overall attitude toward performance enhancing drugs is…

Bad	1	2	3	4	5	6	7	Good
Negative	1	2	3	4	5	6	7	Positive
Unfavorable	1	2	3	4	5	6	7	Favorable

3.4. Independent variables and manipulation checks

The scenario begins with the following: "Imagine you read about a famous athlete. Their career has been a glorious one with many awards, a high paying salary and lots of endorsements. However, recently they are under suspicion of using performance enhancing drugs and a board is reviewing the evidence. No official notice has been given and so no fines have been given. Interestingly, the team and the sponsors are standing by the athlete in an uncharacteristic show of support given the concerns related to this situation."

This scenario establishes a baseline of fame and fortune, creating a context that controls for the influence of other reference groups, such as teammates, team management, and sponsors. Only media coverage (lower and supportive versus higher and unsupportive) is manipulated to operationalize embarrassment. However, in time, all reference groups could be affected if suspicions are verified and revealed by the media. Health consequences are operationalized as minor vs. major and possible death.

Table 1. Independent Variable Manipulation

		Media Coverage	
		High	Low
Health Effects	High	Major health concerns and possible death within the next 10 years; unsupportive, major media coverage; potential relinquishment of a recently received award	Major health concerns and possible death within the next 10 years; supportive, minor media coverage; recently received an award
	Low	Minor health effect; unsupportive, major media coverage; potential relinquishment of a recently received award	Minor health effects; supportive, minor media coverage; recently received an award

Manipulation check - health:

In this scenario the health ramifications of the performance enhancing drugs were…

Not Serious 1 2 3 4 5 6 7 Serious

Manipulation check - embarrassment:

In this scenario the career embarrassment level related to the use of the performance enhancing drugs was… Low 1 2 3 4 5 6 7 High

4. Results

Manipulation checks for Health ($M_{low} = 2.38$, $M_{high} = 5.43$) ($p<.05$) and for Embarrassment ($M_{low} = 2.23$, $M_{high} = 5.28$) ($p<.05$) were both successful. The reliability for attitude had a Cronbach Alpha of (.983) for the three items. The overall model was not significant ($F_{3, 116}=2.5$, $p=.061$), and the adjusted R-Squared = .037. The picture in the scenario, (consistent across conditions) showing the transition after PED usage of a young athlete into a more physically powerful older athlete, showed the benefits of PED usage and was positively correlated to respondent attitudes toward PEDs ($r=.367$, $p=.0001$). This suggests that the picture provided in the scenario was salient and the athletes were aware of the benefit of PEDs, providing a consistent knowledge benchmark for the manipulations.

Table 2. ANOVA Table

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	18.566	3	6.189	2.520	.061
Intercept	1326.675	1	1326.675	540.226	.000
Health	0.675	1	0.675	0.275	.601
Embarrass	17.890	1	17.890	7.285	.008
Health * Embarrass	0.001	1	0.001	0.000	.985
Error	284.870	116	2.456		
Total	1630.111	120			
Corrected Total	303.436	119			

R Squared = .061 (Adjusted R Squared = .037)

The higher the risk of negative health effects, the more negative the athlete's attitude towards performance enhancing drugs will be (H1) was not supported (M Low =3.25, M High =3.4; $F_{1, 116}= 0.28$, $p> .05$). The greater the level of unsupportive media coverage (versus lower and supportive media coverage) revealing an athlete's PED use, the more negative the attitude an athlete will have towards performance enhancing drugs (H2) was supported (M Low =2.94, M High =3.71; $F_{1, 116}= 7.29$, $p = .008$). However, under unsupportive, higher media coverage scenarios (versus lower, supportive media coverage), major health effects result in a more negative change in the athlete's attitude toward performance enhancing drugs, over minor health effects, compared to those found under low media coverage scenarios (H3) was not supported ($F_{1, 116}$= 0.00, p >.05).

5. Conclusions, contributions, limitations, and future research

The most effective communications will feature content that is attended, salient, and a call to action. This study juxtaposed embarrassment (perceived social cost) and serious health failings (perceived health cost) as possible outcomes of the use of PEDs. When asked directly, 44.2% said health cost was important (46.7% if death was a possibility) and 36.7% indicated embarrassment was a factor, with no correlation with their attitude (p <.05). Yet the causal design of this study revealed that only the perceived social cost of embarrassment significantly affected attitudes toward PEDs. This implies that either amateur athletes do not wish to admit the importance of embarrassment, or they were not conscious of its influence on them. Either way, this suggests that marketing communications positioned to emphasize the potential embarrassment (perceived social cost) of being caught may be an effective deterrent to PED usage.

Arguably, amateur athletes are assessing probabilities of perceived costs and yet, the fallacy of their own perceived invincibility may be why dangers to their health are not perceived as personally relevant. Athletes indicated that despite what they read, 14.5% did not "believe anything adverse (health effects) would happen to" them, which in turn was positively correlated with attitudes toward PEDs (r=.425, p=.0001). These beliefs imply their motivations, which in turn may be influenced by reference groups. This is consistent with the TRA. For example, amateur athletes indicated that if they knew someone who uses PEDs, the more positive their attitudes towards PED usage (r=.41, p=.0001). They also indicated that the opinions of others (42.5%) were important in their decision to use PEDs and a substantial number of these athletes indicated that someone they know uses PEDs (36.6%). Further, the success of professional athletes using PEDs was influential (24.2%), and yet, seeing the embarrassment of professional athletes on TV only discouraged 37.5% of the respondents.

Interestingly, the legality of PED usage only concerned 40.8% of the sample, yet only 30.8% indicated that the fear of getting caught would not deter them, and 25.8% never felt they would get caught. Even so, 20% felt embarrassment was irrelevant, as making money was their main goal, with only 42.5% indicating concern about being remembered as a cheat. This suggests that when developing marketing communications to deter PED use, arguably the focus of the message should be on the chances of getting caught by the governing bodies, the potential reaction of the fans, the legacy of being stripped of awards, and the almost tangible aspect of the damage to the user's social image.

Inherent in any experiment, there are certain limitations. As experimental control increases, the complexity of reality may be diminished or discounted. Despite this, these scenarios did elicit a reaction, indicating that respondents were imagining themselves in such a situation. Also, given not all respondents may have reflected on PEDs previously, a set of images were provided to create a connection between PEDs and increased physical ability. However, these pictures focused on one sport by showing a baseball player (a sport with PED use widely publicized). Though this sport was not intended to be the basis for all respondent imaginings, it may have been given unintended weight as a context for their responses. Finally, the amateur athletes sampled were from a university and so represent a more educated sample of athletes.

Future research should explore in more depth the motives behind these choices and whether they differ by types of sports (football vs soccer vs swimming vs baseball vs track and field), or categories of sports (individual vs teams), or level of expertise (amateur vs professional) (Haugen, 2004). Also, it would be interesting to explore further the athlete's rationale for unethical behavior.

That is, do they use PEDs because they do not see PEDs as cheating, or do they rationalize that everyone else is also doing it? Or do PED users accept the reality that they are cheating, but feel they must do so to remain competitive? Our results indicated that attitudes toward using PEDs were positively correlated with the belief that "you have to use PEDs just to remain competitive in sports" (r=.283, p=.002). Finally, do athletes only consider the probabilities of being caught (31% said it was a major concern, 24% were neutral, and 45% were unconcerned), rather than just the seriousness of the outcomes as was this study's focus?

6. References

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50, 179-211.

Bloodworth, A., McNamee, M. (2010). Clean Olympians? Doping and Anti-doping: the view of young British athletes. International Journal of Drug Policy, 21, 276-282.

Chan, D., Dimmock, J., Donovan, R., Hagger, M., Hardcastle, S., and Lentillon-Kaestner, V. (2014). Athletes' beliefs about attitudes towards taking banned performance-enhancing substances: a qualitative study. Sport, Exercise, and Performance Psychology, 3(4), 241-257.

CNN. (2019). Performance enhancing drugs in sports fast facts. CNN. Retrieved from https://www.cnn.com/2013/06/06/us/performance-enhancing-drugs-in-sports-fast-facts/index.html

Connor, J. (2009). Towards a sociology of drugs in sport. Sport in Society, 12(3), 327-328.

Grover, K. & Miller, C. (2014). Effects of mortality salience and perceived vulnerability on HIV testing intentions and behaviour, Psychology & Health, 29(4), 475-490.

Haugen, K. (2004). The performance-enhancing drug game. Journal of Sports Economics, 5(1), 67-86.

Heitner, D. (2015). Sports industry to reach $73.5 billion by 2019. Forbes. Retrieved from https://www.forbes.com/sites/darrenheitner/2015/10/19/sports-industry-to-reach-73-5-billion-by-2019/#3e118cbc1b4b

Howe, L., & Krosnick, J. (2017). Attitude strength. Annual Review of Psychology, 68(3), 327-51.

Jalleh, G., Donovan, R., & Jobling, I. (2014). Predicting attitude towards performance enhancing substance use: A comprehensive test of the Sport Drug Control Model with elite Australian athletes. Journal of Science and Medicine in Sport, 17, 574-579.

Jimenez, A. (2018). The big business of youth sports. National Center for Business Journalism. Retrieved from https://businessjournalism.org/2018/12/business-of-youth-sports/.

Kelly, M., Fukushima, M., Spivak, A., & Payne, D. (2009). Deterrence theory and the role of shame in projected offending of college students against a ban on alcohol. Journal of Drug Education, 39(4), 419-37.

Kirby, K., Moran, A., & Guerin, S. (2011). A qualitative analysis of the experiences of elite athletes who have admitted to doping for performance enhancement. International Journal of Sport Policy and Politics, 3(2), 205-224.

Kisaalita, A, & Robinson, M. (2014). Attitudes and motivations of competitive cyclists regarding use of banned and legal performance enhancers. Journal of Sports and Medicine, 13, 44-50.

Loughran, T., Paternoster, R., Chalfin, A. & Wilson, T. (2016). Criminology, 54(1), 86-112.

Mazanov, J. (2009). Towards a social science of drugs in sport. Sport in Society: Cultures, Commerce, Media, Politics, 12(3), 423-430.

Momaya, A., Fwal, M, & Estes, R. (2015). Performance-enhancing substances in sports: a review of the literature. Sports Medicine, 45(4),517-31.

Murphy, J., & Dennhardt, A. (2016). The behavioral economics of young adult substance abuse. Preventative Medicine, 92, 24-30.

Ntoumanis, N., Ng, J., Barkoukis, V., & Backhouse, S. (2014). Personal and psychosocial predictors of doping use in physical activity settings: a meta-analysis. Sports Med, 44, 1603-1624.

Quirk, F. (2009) Health psychology and drugs in sport. Sport in Society: Cultures, Commerce, Media, Politics, 12(3), 375-393.

Overbye, M., Elbe, A., Knudsen, M., & Pfister, G. (2015). Athletes' perceptions of anti-doping sanctions: the ban from sport versus social, financial, and self-imposed sanctions. Sport in Society, 18(3), 364-384.

Pope, H., Wood, R., Rogol, A., Nyberg, F., Bowers, L., & Bhasin, S. (2014) Adverse health consequences of performance-enhancing drugs: an Endocrine Society scientific statement. Endocrine Review, 35(3), 341-375.

Savulescu, J., Foddy, B., & Clayton, M. (2004). Why we should allow performance enhancing drugs in sport. British Journal of Sports Medicine, 38, 666-670.

Strelan, P., & Boeckmann, R. (2006). Why drug testing in elite sport does not work: perceptual deterrence theory and the role of personal moral beliefs. Journal of Applied Social Psychology, 36(12), 2909-2934.

Ungerleider, S. (1992). Visions of victory. Psychology Today, 25, 45–52.

Wiefferink, C., Detmar, S., Coumans, B., Vogels, T., & Paulussen, T. (2008). Social psychological determinants of the use of performance-enhancing drugs by gym users. Health Education Research, 23(1), 70-80.

World Anti-Doping Agency. (2015). The world anti-doping code. Retrieved from https://www.wada-ama.org/sites/default/files/resources/files/wada-2015-world-anti-doping-code.pdf.

World Anti-Doping Agency. (2009). Dangers of doping: get the facts. Retrieved from https://www.wada-ama.org/en/resources/education-and-prevention/dangers-of-doping-get-the-facts.

An Overview of Very High Cycle Fatigue Behavior of Additively Manufactured Ti-6Al-4V

Palmer Frye[1]
Jutima Simsiriwong[1]
[1]*School of Engineering, College of Computing, Engineering & Construction*
University of North Florida
j.simsiriwong@unf.edu

Abstract

This paper presents a brief review on the current state of knowledge of the very high cycle fatigue (VHCF) behavior of metallic parts fabricated using Additive Manufacturing (AM) processes. It has been shown that AM has significant potential to replace traditional manufacturing methods that impose geometric limitations to designs. Powder-based metallic AM methods allow for precise layer-wise processing of complex net-shape parts without the use of special tooling or molds. Among various metals commonly used in AM processes, titanium (Ti) 6Al-4V alloy is currently of great interest especially in aerospace applications that contain parts with complex geometries (i.e., turbine blades in jet engines). To safely adapt AM Ti-6Al-4V parts in these applications, their mechanical properties and fatigue behavior must be understood. Various studies have identified AM process-induced defects (i.e., entrapped gas pores, lack of fusion defects between build layers, etc.) to be the main cause of fatigue failure of AM Ti-6Al-4V parts. However, there are limited studies relating to the effects of these defects on the behavior of AM metals in the VHCF regime (beyond 107 cycles). Knowledge of the VHCF behavior of AM Ti-6Al-4V is needed for the aforementioned applications due to the high loading frequencies and long service lives required for these parts.

1. Introduction

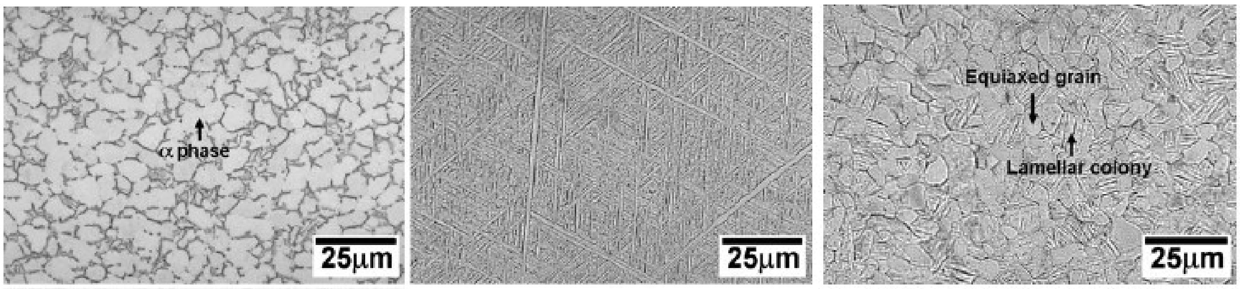
Metallic powder-based additive manufacturing (AM) methods, such as Laser Beam-Powder Bed Fusion (LB-PBF) and Direct Laser Deposition (DLD), can form high density metallic parts in a layer-wise fashion (Xu et al., 2015). Compared to traditional metal manufacturing processes, AM does not require special tooling or molds, which allows for affordable low volume production of functional parts. AM technologies simply require a computer aided design (CAD) model of a part and the metal powder that is to be used. It is expected that the aerospace industry could greatly benefit from metallic AM as it would allow for the production of part geometries that are too difficult or otherwise impossible to create using the traditional manufacturing processes. In particular, the ease of manufacturability of the turbine and compressor blades in jet engines would greatly increase (Gunther et al., 2017).

Titanium (Ti) 6Al-4V is an alloy that is commonly used in the aerospace industry and is also one of the many functionally graded materials available for use in metallic AM. Ti-6Al-4V has many favorable properties for aerospace applications, such as a higher strength to weight ratio as compared to ferrous materials, as well as corrosion and creep resistance (Murr et al., 2008, Heinz & Eifler, 2016). For AM of critical components to surpass those fabricated using traditional manufacturing methods, it is imperative that all material properties, especially fatigue behavior, are fully understood. Since structural fatigue failure accounts for up to 90% of all mechanical failures, extensive material testing and modeling are a necessary benchmark for engineers to properly design parts to withstand cyclic loading (Stephens & Fuchs, 2001). Low and high cycle fatigue (LCF and HCF) behavior studies have already been carried out for AM Ti-6Al-4V; however, critical engine components such as turbine or compressor blades in jet engines require the knowledge of very high cycle fatigue (VHCF) performance which accounts for fatigue lives greater than 107 cycles (Gunther et al., 2017). These critical parts are subjected to very high loading frequencies greater than 1 kHz and require a service life of greater than 109 cycles (Gunther et al., 2017).

Unlike the mechanical properties of conventionally processed metals, those of Ti-6Al-4V parts produced by powder based metallic AM are influenced by many factors, most importantly, process induced defects, residual stresses, and unique microstructures (du Plessis et al., 2018, Gunther et al., 2017, Seifi et al., 2016). Process induced defects in the form of lack of fusion voids from un-melted or un-fused powder particles or pores from entrapped gasses can be detrimental on the mechanical performance (Murr et al., 2008, Seifi et al., 2015, Tammas-Williams et al., 2014). Under cyclic loading, these defects act as stress risers and are common crack initiation sites (Leuders et al., 2012, Gunther et al., 2017). In addition to the high possibility of defects, the rapid heating and cooling associated with metal AM can introduce a variable distribution of residual stresses throughout parts which may have adverse effects on fatigue resistance (Gunther et al., 2017). Various researchers have investigated the effects of post build hot isostatic pressing (HIP), stress relief heat treatment, and manipulation of build parameters as ways to improve mechanical performance (Gunther et al., 2017, Ali et al., 2017). Still, a detailed understanding of these effects is lacking.

2. Effects of Microstructure

A material's microstructure can greatly influence its various mechanical properties (Stephens & Fuchs, 2001). For example, the yield strength of a material is directly related to the size and orientation of the grains within its microstructure. The boundaries between different grains serve as roadblocks for slip; therefore, larger grained metals are more ductile than their smaller grained counterparts (Meyers & Chawla, 2001). Grain size is also known to affect fatigue behavior of Ti-6Al-4V in a similar manner: as the grains become coarser, fatigue resistance decreases (Everaerts et al., 2016, Heinz & Eifler, 2016). Ti-6Al-4V is a two-phase alloy consisting of a hexagonal close-packed alpha phase and a body centered cubic beta phase. Varying thermo-mechanical processing methods plays a critical role in the formation of different microstructures and phase contents in commercially available titanium. The three most common microstructures of Ti-6Al-4V are equiaxed, lamellar, and bimodal (Park et al., 2011). The temperature at which titanium is processed determines the resulting microstructure. Titanium processed above the beta-transus temperature results in lamellar titanium shown in Figure 1b. Lamellar titanium microstructures contain a mixture of long grains of alpha and beta phase ordered in a neat, layered fashion (Meyers & Chawla, 2001). Titanium processed below the beta-transus results in equiaxed and bimodal microstructures, shown in Figures 1(a) and (1c). Equiaxed microstructures consist of primarily large alpha phase crystals, with small amounts of randomly sized and oriented beta crystals (Park et al., 2011). Bimodal microstructures are simply a combination of equiaxed and lamellar titanium microstructures. The material processing discussed in this paper results in microstructures that are somewhat unique due to the inconsistent thermal history of parts manufactured by powder-based AM methods (Galiando-Fernandez et al., 2018).



(a) (b) (c)

Figure 1. Common microstructures of commercially available wrought Ti-6Al-4V: (a) equiaxed, (b) lamellar, and (c) bimodal (Park et al., 2011)

The fabrication process of AM Ti-6Al-4V involves rapid heating and cooling of the metal, which induces a unique microstructure formation (Galiando-Fernandez et al., 2018). The electron backscatter diffraction (EBSD) maps in Figure 2 represent the typical microstructural configurations of Ti-6Al-4V fabricated using LB-PBF and DLD processes. The microstructures shown in Figure 2 are primarily alpha titanium with small amounts of beta titanium scattered between alpha grains. However, it is suggested that the rapid heating with a laser followed by rapid cooling causes a martensitic transformation of alpha titanium, whereas the microstructures of wrought Ti-6Al-4V in Figure 1 are formed through controlled beta-to-alpha heating and cooling (Xu et al., 2015). The EBSD mapping software used in Figure 2 is unable to distinguish between alpha titanium and martensitic alpha prime titanium. The LB-PBF strategy results in a lamellar microstructure of mostly fine columnar alpha-titanium with small amounts of beta-titanium, seen in Figure 2(a). Post build heat treatment of LB-PBF specimens causes a coarsening of the alpha grains, seen in Figure 2(b). These two batches of specimens were tested at both high and low frequencies. Figure 2(c) shows the microstructure of a test specimen that was produced by DLD and has a much finer grained microstructure than any of the test specimens produced by LB-PBF.

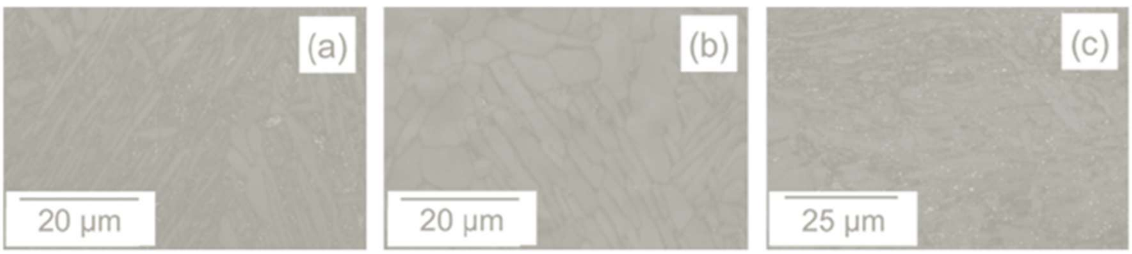


Figure 2. EBSD maps of various microstructures produced by additive manufacturing (a) as build LB-PBF (b) heat treated LB-PBF and (c) DLD (Leuders et al., 2012)

In the VHCF regime, AM Ti-6Al-4V experiences exclusively sub-surface fatigue failure (Gunther et al., 2017). The use of HIP as a post process means of decreasing porosity in AM metallic parts has been proven to improve fatigue performance (Masuo et al., 2017, Gunther et al., 2017). The HIP process leads to a drastic decrease in porosity, which makes test specimens more likely to experience fatigue failure as a result of sub-surface crack initiation at microstructural anomalies or interfaces between alpha and beta titanium (Gunther et al., 2017). In addition, entrapped gas pores have a detrimental effect on VHCF performance of AM titanium, which will be discussed in the next section. Gunther et al. (2017) determined that cracks in HIPed specimens tended to originate from smooth features or facets in the microstructure that are difficult to identify, as compared to non-HIPed specimens with much greater amounts of porosity. The two dominating crack initiation sites after

HIP processing are single alpha-phase facets or clusters of alpha-phase facets, respectively (Gunther et al., 2017). Pores in the microstructure were not fully removed from HIP, but no internal cracks were found to have propagated from pores. Fatigue loading can clearly identify the weakest link in microstructure (Gunther et al., 2017). The current results reveal that only a decrease of the size of the alpha-phase clusters could further improve the fatigue life of additively manufactured and HIPed Ti-6Al-4V (Gunther et al., 2017).

Microstructures of titanium parts built using laser AM can vary for many reasons: powder batches from material suppliers can contain inconsistent chemistries, different operating platforms, differing laser parameters, etc., which can all affect microstructural development in AM metallic parts. In addition to variable build parameters, post processing methods, such as heat treatment and machining, can result in parts with inconsistent mechanical performance. Therefore, developing a microstructure sensitive fatigue life model for AM metals seems to be a necessity as AM parts are increasingly adopted in various applications (Sterling et al., 2016).

3. Effects of Defects

Metallic parts that are produced using powder-based AM methods contain defects that are inherent to the process (Seifi et al., 2015). These defects include a lack of fusion of melted or partially melted metal powder between build layers, entrapped gas pores, voids, and poor as-built surface finishes (Seifi et al., 2015, du Plessis et al., 2018). All of these defects are not ideal for parts that experience cyclic loading, as they are likely sites for fatigue crack initiation (Wycisk et al., 2015). Most of these defects can be attributed to non-ideal process parameters and build conditions, such as incorrect laser energy, laser scan speed, humidity of the build chamber, and metal particle size distribution (du Plessis et al., 2018). It is important that the effects of each defect type on mechanical properties and fatigue performance are fully understood because of the random nature of defect formation in AM.

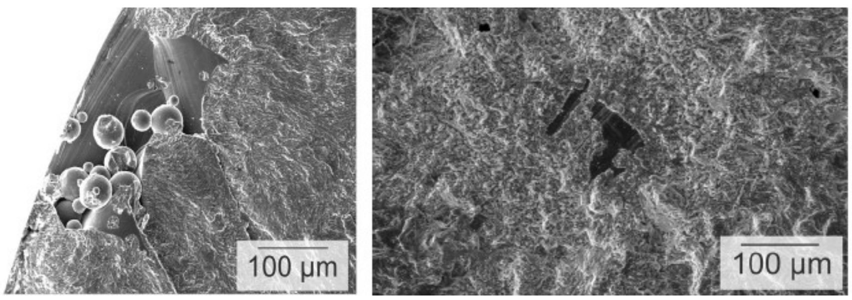
Often during the build process, metallic powder particles are not fully melted and bonded to previously deposit metal layers (Seifi et al., 2015). This type of defect is called lack of fusion (LOF) and can happen on both part surfaces and sub-surfaces between build layers. LOF defects are usually a result of insufficient overlap of successive melt pools between layers (du Plessis et al., 2018).

High amounts of entrapped gas pores and voids can also exist in AM metallic parts. It was suggested by Tillmann et al. (2017) that the rapid heating and cooling of metals used in LB-PBF does not allow enough time for outgas to occur, which results in pockets of entrapped gasses within parts. In some cases, empty spaces between metal powder particles are not filled during the melting and solidification of build layers, which leads to the presence of micro-voids. The Presence of voids and entrapped gasses reduces the relative density of parts and can cause localized reductions in the strength of the material (Pan et al., 2017). Researches such as Gunther et al. (2017) have implemented HIP as a means of closing the gaps between entrapped gas pores or micro-voids.

Lastly, AM parts may experience noticeable flaws on part surfaces, such as artificial notches, stair-stepping, and separation between build layers. These surface flaws act as stress concentrations, which can be detrimental to fatigue and mechanical properties (du Plessis et al., 2018). These defects are characteristic of parts built in a powder bed, as unmelted powder particles bond the part surfaces while the laser-melted metal solidifies. Post process machining and polishing of part surfaces are the only possible means of minimizing the detrimental effects of surface roughness on the fatigue behavior of AM parts. However, it is difficult to assess the exact effects of the as-built surface quality of AM metallic parts mechanical behavior. As-built AM specimens are rarely built to the same dimensional tolerances of test specimens that have been machined. As-built specimens are still expected to fail due to a large amount of pores and internal defects (Edwards et al., 2014).

Fatigue performance of AM Ti-6Al-4V is greatly influenced by the aforementioned process induced defects (Wycisk et al., 2015). In wrought titanium, crack initiation is mostly a function of the microstructure; phase interphases at part surfaces or at part sub-surfaces are likely crack initiation sites since wrought titanium is typically of much lower porosity than AM titanium (Crupi et al., 2016). In AM, the dominating factors influencing crack growth behavior are porosity and LOF defect population. The subsequent fatigue cracks that occur as a result of these defects have been found to initiate from both surface and sub-surface defects (Gunther et al., 2017, Leuders et al., 2012). Post build treatment by HIP removes many of the detrimental process induced defects and significantly improves fatigue performance (Wycisk et al., 2015). AM metallic parts that lack process inherent defects can experience fatigue failures similar to the wrought material, in which cracks are initiated by the weakest link in the microstructure (Gunther et al., 2017, Wycisk et al., 2015).

In the LCF and HCF regimes, AM Ti-6Al-4V experiences fatigue failure from surface and sub-surface fatigue cracks (Gunther et al., 2017). The internal defects that are most likely to initiate fatigue cracks are LOF defects (i.e. portions of partially melted or un-melted sections of metal powder). However, at higher stress amplitudes, fatigue failure is also likely to occur from sharp notches unmelted powder on part surfaces. An example of a surface fatigue failure site can be seen in Figure 3(a). It is clear from Figure 3(a) that the apparent discontinuity on the surface is the crack initiating defect. As fatigue life exceeds 107 cycles, AM Ti-6Al-4V experiences an apparent shift from surface to sub-surface fatigue failure. LOF defects such as those shown in Figure 3(a) are also likely to manifest beneath part surfaces. These defects lack uniformity in shape and size and are dispersed randomly throughout the microstructure. The fracture surface in Figure 3(b) is that of a specimen which failed as a result of a large, sub-surface LOF defect.



(a) **(b)**

Figure 3. LOF defects responsible for fatigue crack initiation in Ti-6Al-4V specimens with (a) $N_f = 1.56 \times 10^5$ cycles and (b) $N_f = 1.17 \times 10^7$ cycles (Gunther et al., 2017)

Entrapped gas pores and micro-voids in test samples had the most significant influence on fatigue performance (Gunther et al., 2017, Leuders et al., 2012). Gunther et al. (2017) compared two sets of specimens built under the same process parameters, however, one set of specimens was HIPed and the other set was only stress relieved. Figure 4 compares the stress-life performance between the two sets of specimens. Clearly, the HIP process significantly reduced the porosity of the specimens, and as a result drastically improved fatigue performance as compared to the non-HIPed specimens (Gunther et al., 2017). However, even the HIPed specimens contain some remnant porosity and voids, which shows that the HIP process only minimizes the presence of pores only. Micro-voids cannot be removed by any post processing techniques. This same result has also been observed in other studies (Masuo et al., 2017).

The cost and environmental effects associated with HIP may burden the benefits of AM. The presented studies have shown that HIP is the most effective post-process method to increase fatigue resistance of AM titanium parts. It would be more advantageous to the manufacturer if parts did not require HIP or any post build heat treatments to be used in cyclically loaded applications. Kasperovich and Haussman (2015) successfully optimized laser scan parameters as a means of decreasing remnant porosity. Optimal build parameters were able to decrease porosity to below %0.077 of the total part volume, subsequent HIP only further decreased this value (Kasperovich and Haussman, 2015).

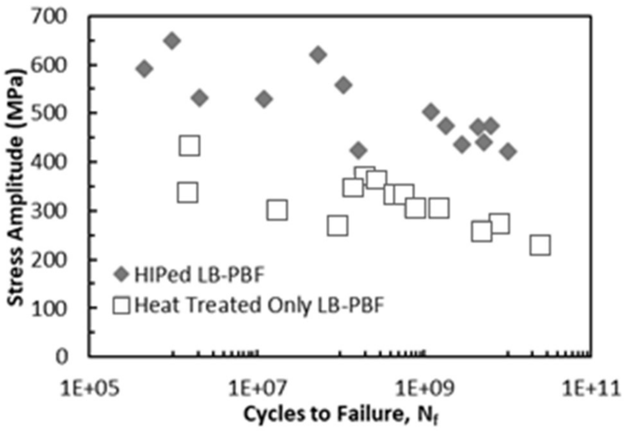


Figure 4. Comparative stress-life (S-N) data of LP-PBF Ti-6Al-4V specimens with and without HIP treatment (Gunther et al., 2017)

Based on the results of the presented studies, it is evident that AM Ti-6Al-4V can be adequately used in structural applications, after a series of post-build treatments. HIP treatment of as-built AM metallic parts is a necessary first step for diminishing the presence of process inherent defects, which are the most common sites for cracks to initiate. In order to advance AM of metals like titanium, continuous improvement of AM technologies, post processing methods, characterization, and mechanical testing are required.

4. Effects of Residual Stresses

AM allows for manipulation of several process control parameters such as laser scan speed, build platform temperature, cooling rates, and vacuum pressure of the chamber (Ahmad et al., 2018). It is known that all these parameters affect the microstructural development of the part, and coincidentally affect mechanical properties and fatigue performance (Seifi et al., 2015, Gunther et al., 2017). Microstructural development of AM metallic parts is primarily a function of build temperature. Wook Na et al. (2018) determined that by increasing the power of the laser in LB-PBF Ti-6Al-4V leads to a higher Vickers hardness of the part. Wook Na et al. (2018) concluded that by increasing the laser power, the overall temperature of the melt pool increased. This increase of melt temperature raises the partial pressure of the gasses trapped within the melt, which increases the chance of outgas (Wook Na et al., 2018)

Gunther et al. (Gunther et al., 2017) analyzed the effects of different heat treatment methods on fatigue behavior of AM Ti-6Al-4V. Three batches of test specimens were built using LB-PBF, two of which were heat treated for two hours at 800°C in Argon, and the third batch was HIPed for two hours at 920°C and 1000 bar. Wrought Ti-6Al-4V test specimens that are stress relieved after machining are known to exhibit superior fatigue performance in the VHCF regime, but this is mostly due to microstructural changes imposed by heat treatment (Adams et al., 2016). The rapid heating and cooling of metal that takes place during the AM process induces the detrimental residual stresses effects on fatigue resistance in parts (Ahmad et al., 2018, Li et al., 2018). In general, compressive residual stresses can be beneficial for fatigue performance, especially in retarding crack initiation and propagation (Stephens & Fuchs et al., 2001).

Distribution of residual stresses in AM metallic parts can be correlated to the build layer orientation. It was found that tensile residual stresses are of the highest magnitude along the build direction and on outer surfaces of LB-PBF parts, while compressive residual stresses are more concentrated in the part center (Ali, 2017). Heat treatment is not the only way to reduce residual stresses. In situ observation of residual stress development in AM Ti-6Al-4V shows that pre-heating the build plate of an LB-PBF machine, and subsequently decreasing the cooling rate to a value much lower than if no pre-heating was applied, significantly increases ductility, which is characteristic of stress relieved parts (Ali et al., 2017).

Residual stress development in parts also varies depending on the AM process utilized. Tensile residual stresses are highest in as-built LB-PBF Ti-6Al-4V test specimens, as compared to specimens built with DLD (Gunther et al., 2017). This is mostly due to the significant difference in cooling rates between the two build processes. DLD build chambers can be held at higher temperatures than LB-PBF machines are capable of. The residual stresses in as-build DLD specimens were found to be very similar in value and distribution to those in LB-PBF specimens. DLD provides an immediate advantage over LB-PBF because superior mechanical properties can be achieved by DLD without the need for post process heat treatment (Gunther et al., 2017).

The effect of residual stresses on fatigue performance of AM Ti-6Al-4V has not yet been studied directly. Siddique et al. (2016) observed both quasi-static and VHCF properties of AM AlSi12 manufacturing using LB-PBF with build platform heating in lieu of post processing techniques such as HIP and heat treatment. The method of build platform heating significantly reduces the temperature difference between the build platform and the metal as it is melted by the laser. This lowers the cooling rate experienced by the metal, which consequently lowers the ultimate tensile strength and yield strength due to coarsening of the microstructure (Siddique et al., 2016). Figure 5 is the resulting S-N curve of this study (Siddique et al., 2016). Build platform heating is an effective means for enhancing VHCF performance of LB-PBF parts without additional post processing.

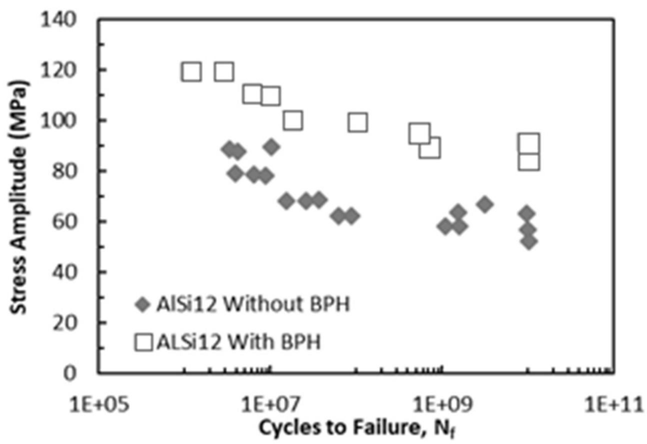


Figure 5. Fatigue life curve of AM AlSi12 up to VHCF with and without build platform heating (BPH) (Siddique, 2016)

5. Conclusions

Fatigue performance of additively manufactured metallic parts shows a strong dependence on microstructural characteristics (i.e., grain size and orientation), as well as phase composition in multiphase materials. In most cases, the presence of surface flaws and internal defects are the two distinct characteristics that lead to fatigue failure in AM metallic parts. In general, it can be concluded that VHCF failure is exclusively due to cracks initiation from process induced surface defects, and that fewer surface cracks can lead to much longer fatigue lives in Ti-6Al-4V specimens. After 106 cycles and up to 109 cycles, the main cause of fatigue failure is the lack of fusion of metal powder or high amounts of porosity. Nonetheless, VHCF performance can be significantly improved by implementing post process HIP to minimized remnant porosity.

Fatigue lives of AM specimens seem to follow the trend observed in the wrought counterparts: as the applied stress decreases, the dominating site for fatigue failure shifts from surface failure to the sub-surface failure. Moreover, the relationship between residual stresses and VHCF of AM metallic parts is not fully understood due to very limited studies. Stress relief of AM metallic parts was found to increase the fatigue strength due to the coarsening of the microstructure. Build plate heating in the LB-PBF process could be utilized as an alternative to post process stress relief as it does not affect the build time, but it has been shown to increase the VHCF resistance over the specimens fabricated on the non-heated build plate.

While the studies presented in this overview paper show a promising future for AM of titanium for structural parts, there is still a significant lack of reliable fatigue data, which means that immediate adaptation of AM as a production method may not yet be applicable. Until Ti-6Al-4V is adequately characterized for its mechanical behaviors and performance, and AM build and test parameters are streamlined by global standards, it may remain in low stress, low risk applications.

6. References

Ahmad, B., Veen, S. O., Fitzpatrick, M. E., & Guo, H. (2018). Residual stress evaluation in selective-laser-melting additively manufactured titanium (Ti-6Al-4V) and inconel 718 using the contour method and numerical simulation. Additive Manufacturing,22, 571-582.

Ali, H., Ma, L., Ghadbeigi, H., & Mumtaz, K. (2017). In-situ residual stress reduction, martensitic decomposition and mechanical properties enhancement through high temperature powder bed pre-heating of Selective Laser Melted Ti6Al4V. Materials Science and Engineering: A,695, 211-220.

Bača, A., Konečná, R., & Nicoletto, G. (2017). Influence of the Direct Metal Laser Sintering Process on the Fatigue Behavior of the Ti6Al4V Alloy. Materials Science Forum,891, 317-321.

Crupi, V., Epasto, G., Guglielmino, E., & Squillace, A. (2017). Influence of microstructure [alpha beta and beta] on very high cycle fatigue behaviour of Ti-6Al-4V alloy. International Journal of Fatigue,95, 64-75.

Edwards, P., & Ramulu, M. (2014). Fatigue performance evaluation of selective laser melted Ti–6Al–4V. Materials Science and Engineering: A, 598, 327-337.

Everaerts, J., Verlinden, B., & Wevers, M. (2016). The influence of the alpha grain size on internal fatigue crack initiation in drawn Ti-6Al-4V wires. Procedia Structural Integrity,2, 1055-1062.

Galindo-Fernández, M., Mumtaz, K., Rivera-Díaz-Del-Castillo, P., Galindo-Nava, E., & Ghadbeigi, H. (2018). A microstructure sensitive model for deformation of Ti-6Al-4V describing Cast-and-Wrought and Additive Manufacturing morphologies. Materials & Design,160, 350-362.

Günther, J., Krewerth, D., Lippmann, T., Leuders, S., Tröster, T., Weidner, A., Biermann, H., &Niendorf, T. (2017). Fatigue life of additively manufactured Ti–6Al–4V in the very high cycle fatigue regime. International Journal of Fatigue,94, 236-245.

Günther, J., Leuders, S., Koppa, P., Tröster, T., Henkel, S., Biermann, H., & Niendorf, T. (2018). On the effect of internal channels and surface roughness on the high-cycle fatigue performance of Ti-6Al-4V processed by

SLM. Materials & Design,143, 1-11.

Heinz, S., & Eifler, D. (2016). Crack initiation mechanisms of Ti6Al4V in the very high cycle fatigue regime. International Journal of Fatigue,93, 301-308.

Kasperovich, G., Hausmann, J., (2015) Improvement of fatigue resistance and ductility of TiAl6V4 processed by selective laser melting. J Materials Process Technology; 220:202–214.

Leuders, S., Thöne, M., Riemer, A., Niendorf, T., Tröster, T., Richard, H., & Maier, H. (2013). On the mechanical behaviour of titanium alloy TiAl6V4 manufactured by selective laser melting: Fatigue resistance and crack growth performance. International Journal of Fatigue,48, 300-307.

Li, C., Liu, Z., Fang, X., & Guo, Y. (2018). Residual Stress in Metal Additive Manufacturing. Procedia CIRP,71, 348-353.

Masuo, H., Tanaka, Y., Morokoshi, S., Yagura, H., Uchida, T., Yamamoto, Y., & Murakami, Y. (2018). Influence of defects, surface roughness and HIP on the fatigue strength of Ti-6Al-4V manufactured by additive manufacturing. International Journal of Fatigue,117, 163-179.

Masuo, H., Tanaka, Y., Morokoshi, S., Yagura, H., Uchida, T., Yamamoto, Y., & Murakami, Y. (2018). Influence of defects, surface roughness and HIP on the fatigue strength of Ti-6Al-4V manufactured by additive manufacturing. International Journal of Fatigue,117, 163-179.

Murr, L., Quinones, S., Gaytan, S., Lopez, M., Rodela, A., Martinez, E., Hernandez, D., Martinez, E., Medina, F., & Wicker, R. (2009). Microstructure and mechanical behavior of Ti–6Al–4V produced by rapid-layer manufacturing, for biomedical applications. Journal of the Mechanical Behavior of Biomedical Materials,2(1), 20-32.

Meyers, M., Chawla, K., (1998) Mechanical Behavior of Materials. New York: Cambridge

Na, T., Kim, W. R., Yang, S., Kwon, O., Park, J. M., Kim, G., Jung, K.H., Lee, C.W., Kim, H. G. (2018). Effect of laser power on oxygen and nitrogen concentration of commercially pure titanium manufactured by selective laser melting. Materials Characterization,143, 110-117.

Pan, X., Su, H., Sun, C., & Hong, Y. (2018). The behavior of crack initiation and early growth in high-cycle and very-high-cycle fatigue regimes for a titanium alloy. International Journal of Fatigue,115, 67-78.

Park, C. H., Son, Y. I., & Lee, C. S. (2011). Constitutive analysis of compressive deformation behavior of ELI-grade Ti–6Al–4V with different microstructures. Journal of Materials Science,47(7), 3115-3124.

Persenot, T., Burr, A., Martin, G., Buffiere, J., Dendievel, R., & Maire, E. (2019). Effect of build orientation on the fatigue properties of as-built Electron Beam Melted Ti-6Al-4V alloy. International Journal of Fatigue,118, 65-76.

Plessis, A. D., Yadroitsev, I., Yadroitsava, I., & Roux, S. G. (2018). X-Ray Microcomputed Tomography in Additive Manufacturing: A Review of the Current Technology and Applications. 3D Printing and Additive Manufacturing,5(3), 227-247.

Seifi, M., Christiansen, D., Beuth, J., Harrysson, O., & Lewandowski, J. J. (2016). Process Mapping, Fracture and Fatigue Behavior of Ti-6Al-4V Produced by Ebm Additive Manufacturing. Proceedings of the 13th World Conference on Titanium,1373-1377.

Seifi, M., Salem, A., Satko, D., Shaffer, J., & Lewandowski, J. J. (2017). Defect distribution and microstructure heterogeneity effects on fracture resistance and fatigue behavior of EBM Ti–6Al–4V. International Journal of Fatigue,94, 263-287.

Siddique, S., Imran, M., Wycisk, E., Emmelmann, C., & Walther, F. (2016). Fatigue Assessment of Laser Additive Manufactured AlSi12 Eutectic Alloy in the Very High Cycle Fatigue (VHCF) Range up to 1E9 cycles. Materials Today: Proceedings,3(9), 2853-2860.

Stephens, R. I., & Fuchs, H. O. (2001). Metal fatigue in engineering. New York: Wiley.

Sterling, A. J., Torries, B., Shamsaei, N., Thompson, S. M., & Seely, D. W. (2016). Fatigue behavior and failure mechanisms of direct laser deposited Ti–6Al–4V. Materials Science and Engineering: A,655, 100-112.

Sun, W., Ma, Y., Ai, X., & Li, J. (2018). Effects of the building direction on fatigue crack growth behavior of Ti-6Al-4V manufactured by selective laser melting. Procedia Structural Integrity,13, 1020-1025.

Tammas-Williams, S., Zhao, H., Léonard, F., Derguti, F., Todd, I., & Prangnell, P. (2015). XCT analysis of the influence of melt strategies on defect population in Ti–6Al–4V components manufactured by Selective Electron Beam Melting. Materials Characterization,102, 47-61.

Vayssette, B., Saintier, N., Brugger, C., Elmay, M., & Pessard, E. (2018). Surface roughness of Ti-6Al-4V parts obtained by SLM and EBM: Effect on the High Cycle Fatigue life. Procedia Engineering,213, 89-97.

Wycisk E, Siddique S, Herzog D, Walther F, & Emmelmann C. (2015) Fatigue performance of laser additive manufactured Ti–6Al–4V in very high cycle fatigue (VHCF) regime up to 109 cycles. Front Mater; 2:72.

Xu, W., Sun, S., Elambasseril, J., Liu, Q., Brandt, M., & Qian, M. (2015). Ti-6Al-4V Additively Manufactured by Selective Laser Melting with Superior Mechanical Properties. JOM,67(3), 668-673.

Engineering Disasters: The Role of Engineering versus Management Cumulative Failure Risk Factor

Andrzej J. Gapinski[1]
[1]*The Pennsylvania State University - Fayette*
ajg2@psu.edu

Abstract

The article investigates engineering disasters as failures of either engineering design, project management decisions, or management processes in general. The paper points out that more often than not the failures of engineering endeavors were due to shortcomings of project management and organizational culture irrespective of the area of engineering discipline involved. The cumulative failure risk factor is proposed to assess an overall project failure risk, which can assist in project failure risk assessment and consequently in identifying the shortcomings in an organization.

1. Introduction

The engineering field has delivered numerous marvels and achievements through the millennia in various civilizations. The pace only accelerated in the last two centuries. These achievements not only improved the lives of many societies worldwide, but also delivered breakthrough discoveries and technologies. While nobody questions the engineering achievements of the past, there have been some disasters along the way. The article analyzes engineering disasters and their causes, which may vary from faults in engineering design, faults in the implementation processes, failures in project management with respect to decision making processes, and other factors. The article identifies the human aspect, specifically, the interplay between engineering and management, as the culprit of engineering disasters and project failures.

2. Examples of project failures

From the time of antiquity with its architectural and civil engineering marvels to modern times, failures and project disasters were part of the learning process in human engineering endeavors. Antiquity provides examples of disasters in the areas of design shortcomings, exceeding design specifications, and poor workmanship. In his article, Rogers (2014) lists examples and illustrations of such structural disasters in ancient and medieval times: the Bent Pyramid of Egypt 2600 BCE, the Fidenae Amphitheater collapse in Italy 27AD, the Circus Maximus upper tier collapse, Italy 140 AD, the Beauvais Cathedral collapse, France (1284), and the Rialto Bridge collapse, Italy (1444). The causes of the above-mentioned failures include exceeding carrying weight limits, faults in the design including facts unknown to builders at the time of design, physical phenomena (mechanical resonance), poor workmanship, or overambitious designs by commissioning authorities.

2.1 Examples of Engineering Failures in Modern Times

The practicum of engineering provides many examples of failures in modern times. Some well publicized examples that occurred in the USA are: the Tacoma Narrows Bridge collapse in 1940 in the state of Washington (Tacoma, 1940), the collapse of a cooling tower in 1978 in West Virginia (Willow Island, 1978), the Three Mile Island nuclear power plant disaster in Pennsylvania (Three Mile Island, 1979), and the more recent shuttle Challenger disaster in 1986 (Challenger, 1986). These were dramatic examples of catastrophic failures in civil, nuclear, and aerospace engineering. Specifically, the shortcomings in the understanding of the resonance effect (Tacoma Bridge), rushing through tower construction and evident violation of construction methods (Willow tower), a series of mechanical and human errors leading to nuclear core meltdown (Three Mile Island, 1979), the flaw in the design of the shuttle booster sealing-rings and management's overrule of engineering staff safety concern (Challenger, 1986), respectively, were identified causes of the listed accidents. One may list additional illustrations of industrial and environmental disasters such as the Donora Smog of 1948 in Pennsylvania where air inversion trapped toxic gasses around the industrial town of Donora killed its inhabitants (Koetz, 2018), the Quecreek Mine accident where the close vicinity of the poorly documented abandoned Saxman mine led to barrier penetration and accidental flooding of Quecreek Mine, Pennsylvania (Quecreek, 2013), the Deepwater Horizon (Deepwater Horizon, 2010) Gulf of Mexico oil spill accident due to faulty well design that did not include enough cement, and disasters on foreign soil such as the Chernobyl nuclear power plant disaster in Ukraine, 1986 due to "improper testing of the nuclear reactor" and "lack of safety mechanism on the reactor" (Chernobyl, 1986), or Bhopal, India where a toxic gas leak due to "slack management and deferred maintenance" or outright "sabotage" as claimed by plant owner Union Carbide Corporation, caused the death of thousands of inhabitants in towns adjacent to the plant (Bhopal, 1984).

Love et al. (2011) analyzed failures in structural engineering and construction based on many examples including: the Hyatt Regency Hotel multistory atrium collapse in 1981 (faulty modified design with no needed calculations performed; the design supported only 60% of the minimum load required by building codes as reported by Moncranz and Taylor, 2000), the Charles de Gaulle International Airport in France, 2004 (the vaulted roof collapsed due to exterior temperature swings; insufficient margin for safety in the design), Melbourne's West Gate Bridge 115 meter span collapse in 1970 (structural failure due to structure overweight), or Singapore's New World Hotel collapse in 1986 (faulty concrete composition as reported by Thean et al., 1987).

Workforce (2018) analyzed less dramatic cases than above mentioned ones, the failures in project or product development or product marketing that include: Sony Betamax, Polaroid Instant Home Movies, Apple Lisa, IBM PCjr, DeLorean DMC-12, Ford Edsel, etc. Some of the reasons for failures were: failing to follow up on project relevancy (Sony), failure of staying abreast of current market needs (Polaroid), over promising and under delivering with a product (Apple), low product quality (IBM), production and quality issues (DMC-12), and inadequate speed to market (Edsel).

2.2 Electrical Systems and Information Technology Disasters

Countless electrical devices, equipment, and computer networks provide not only convenient amenities, but also critically important services responsible for the overall efficient functioning of modern infrastructure. Consequently, it is justified to address electrical systems and information technology (IT) failures separately.

Electrical failures include: failures of power/distribution station, failures of transformers (overloading, sudden surges of currents/voltages exceeding circuit breakers specifications), arc-flush accidents (safety precautions not taken, lack of PPE, poor maintenance, etc.), occasional power grid blackouts, etc. Electrical power system failures include smart grid malfunctions, which may increasingly be caused by IT deficiencies (Electrical Disasters, 2018; Transformers, 2018). Electric power blackout of 2003 affecting north-eastern continental USA and Canada was caused by power overloading and software bug in the alarm system (Blackout, 2003).

Rae and Eden (2002), in their paper, discuss failures in engineering projects based on the analysis of twelve projects related to the electrical power generation industry. They pointed out that the greatest impact on project management outcomes were: "project delays, engineering advisors failing in controlling project engineering, and changes in (project) specification."

3. Project disaster defined

Morris and Hough (1987) gave the following description: a project disaster occurs when the answers to the following questions are all negative, as noted by Storm and Savelsbergh (2014):

1. "Project functionality: does the project perform financially, technically, or otherwise in the way expected by the project sponsors?
2. Project management: was the project implemented to budget, on schedule, to technical specifications?
3. Contractors' commercial performance: did those who provided a service for the project benefit commercially?
4. Cancellation: in the event that a project needed to be canceled, was the cancellation made on a reasonable basis and terminated efficiently?"

Peter Hall (1980) defines project disaster as: "any planning process that is perceived by many people to have gone wrong." In establishing the prospects for a project to be successful one has to assess the risks involved. Larson and Gary (2014) write "in the context of projects, risk is an uncertain event or condition that, if it occurs, has a positive or negative effect on project objectives."

Naturally, in the classification of failures, one has to take into account the severity and impact of the accidents. Here, a disaster is understood as a catastrophic failure resulting in major inconveniences, more drastic effects causing fatalities of personnel or a population, or a major degradation of the environment.

The substantial subjectivity in defining what constitutes a project failure or disaster is pointed out by many researchers. Authors such as Hall (1980), Weick and Sutcliffe (2001) stressed that for proper identification of causes of a project failure, it is necessary to understand the whole project's process as a dynamic entity. Consequently, the recent trends in analyses of project failures assign more significance to objective criteria where project failures are assessed from the perspective of meeting the process and design specifications and efficiencies of process inner-working. Organizational propensity for failure is a subject of study in the discipline of risk management (Chapman and Ward, 2003; Heldman, 2005; Crouhy et al., 2005).

4. Project failures causes

The discipline of risk management, according to Gido and Clements (2015), "involves identification, assessments, control, and response to project risks to minimize the likelihood of occurrence and/or potential impact of adverse events on the accomplishment of the project objective." Larson and Gray (2014) provide the following categories of project failures: "technical, external, organizational, and project management".

Many authors have performed analysis of project failures using multi-criteria approaches and provided a classification of causes. Based on an empirical study performed and reported by Storm and Savelsbergh (2014), the causes in descending reported frequency or importance are:

1. "Poor project management.
2. Weak business case and inadequate attention to business needs and goals.
3. Lack of top management support.
4. Lack of attention to the human and organizational aspects.

5. Failure to involve users appropriately.
6. Inappropriate contracts.
7. Inadequate design solutions.
8. Incompetence and lack of experience."

Based on their surveys among project managers, Discenza and Forman (2007) list the following factors causing project failures in descending order of significance:

1. project communications (business value vs technical aspects, customer involvement) – 43% of responding project managers
2. project's process inner-working factors (accountability, consistency in planning & execution, having checkpoints) – 42% of responders
3. the human factor (managing and motivating people, providing needed tools) – 32% of responders.

These causes of failures are listed by many authors. Rae and Eden (2002) in their paper on the electric power generation systems projects conclude that the major contributing factors to project failures were: "project management system, ineffective organizational procedures, or practices." In a Computerworld article, Rosencrance (2007) referred to a CompTIA web poll that stated that poor communication causes most IT project failures. In the report, communication was understood as the ability to "communicate the project objectives, the expected results, and the budget restrictions." Furthermore, the CompTIA' poll listed "insufficient resource planning" and "unrealistic deadlines" as other factors. Although the CompTIA analysis was drawn from IT sector, the conclusions of the analysis are applicable to other areas as well. Many authors classify over-runs in time and costs as project failures, but according to empirical data over-runs are not uncommon. In their analysis of 3500 projects performed all over the world in various industries, Morris and Hough (1987) discovered that cost over-runs are "normal," varying between 40% and 200%, especially in large-scale projects. Consequently, the over-runs either in time or in costs merit a separate subcategory as a failure cause.

5. Engineering vs Management: Does Wearing a Different "Hat" Affect Risk Tolerance? The Role of Groupthink, Communication, and Organizational Culture in Decision Making

Many authors singled out shortcomings in communication as a main contributor to the overall success or failure of any project. Thus, the human factor cannot be underestimated in any business or engineering project endeavor. Management and engineers perform different functions that have different objectives. While engineers' task focuses on delivering a design to meet product specifications, management is concerned with costs and time limits. In addition, management is often the subject of external pressure to meet customer expectations. Shortcuts taken to mitigate external pressure may result in catastrophic outcomes. To address the issue of conflicting demands put on engineering and management, this article analyzes the case of the Challenger Shuttle disaster.

The Challenger disaster was the subject of extensive studies including a report by a government commission (Rogers Commission, 1986) and books written by people directly involved in the project. The analysis presented in the book by McDonald and Hansen (2009) provides a rare insight into inter-organizational dependencies where various interests collide, and safety is sacrificed to meet project deadlines.

The Rogers Commission, according to McDonald and Hansen (2009), "reported four major findings…First, the cause of the accident was frozen rubber O-rings in the SRB [solid rocket booster] joints, which allowed a leak of burning fuel. Second, engineers working at Marshall and at Morton Thiokol in Utah, the SRB contractor, knew that the joint design was dangerous, especially in cold temperatures. Third, Marshall [Space Flight Center in Huntsville Alabama] project managers had known for some time that the joints were hazardous but failed to communicate that understanding to chief Shuttle officials at Johnson Space Center in Houston and to NASA headquarters during preflight reviews in Washington. Fourth, MSFC [Marshall Space Flight Center in Huntsville Alabama] officials botched the last-minute teleconference with Thiokol, held the evening prior to launch. They pressured Thiokol's top managers to overrule their engineers and recommend launch even in the cold weather expected the next morning at Cape Canaveral." Thus, the Morton Thiokol engineering staff was overruled in their decision not to launch. McDonald writes that Roger Boisjoly (a Morton Thiokol leading engineer) "has been charging that NASA management in its MSFC center played 'fast and loose' with astronauts' lives, 'absolutely abdicating their professional responsibility' in pressuring Thiokol to reverse its original recommendation not to launch."

Regarding the deficiencies in communication by NASA, the Rogers Commission (1986) in its report points out shortcomings in communication by NASA Marshall Space Flight Center: "The Commission found that Marshall Space Flight Center project managers, because of a tendency at Marshall to management isolation, failed to provide full and timely information bearing on the safety of flight 51-L to other vital elements of Shuttle program management."

McDonald points out that Morton Thiokol management should also take responsibility for caving to NASA's pressure and reversing its engineering staff recommendation not to launch. Collins Michael (command module pilot for Apollo 11 mission), among others, suggested a "cult of arrogance," writes McDonald, which pervaded NASA and contributed ultimately to the shuttle incident. According to Boisjoly, writes McDonald, "NASA officials were so determined to launch Challenger that the top Shuttle experts forced Thiokol to prove beyond any doubt that it was not safe to do so – when in most flight readiness reviews officials had to prove just the opposite."

Thus, the engineering and management staff were subject to conflicting interests and were positioned on opposite sides of safety and timely deliverables. Organizational culture definitely plays a role in creating an environment prone to an increased probability of failure. In his book, Tompkin (1985) expresses the view that organizations may suffer from "ignorantia affectata" or a cultivated arrogance, and individuals must take responsibility for their own actions" ultimately to minimize or prevent the failures.

In their analysis of the shuttle disaster, Pinkus et al. (1997) point to communication shortcomings, if not breakdowns within NASA: "NASA's top decision makers were never informed of Thiokol's concerns. Marshall (Marshall Space Flight Center) officials chose not to pass this information to their superiors." They conclude that "no doubt the organizational hierarchy of NASA was a considerable impediment for negative information reaching the top of the organization." However, the authors point out that NASA's administration was frequently facing, during multiple launches, objections to launching from "engineers who, like those at Thiokol, advised against the launch, claiming that the shuttle could experience a catastrophic failure."

Pinkus et al. (1997) in their analysis from the ethical consideration point of view note deficiencies in core ethical attributes such as individual and organizational competency (understanding of technical data, lack of understanding/appreciation of statistical data analysis, obligation to seek knowledge even outside an organization, etc.) and responsibility (obligation to voice the concern about safety issues, lack of organizational responsiveness, etc.), violation of "Cicero Creed II" wherein "engineers should understand and characterize the risks associated with technology." Pinkus et al. (1997) claim, based on existing records, that the NASA and Thiokol engineering staff was lacking "statistical knowledge" that "might have been critical in making the O-ring performance decision." To illustrate this they constructed, using a regression technique, a probability estimation model of joint failure to fit the pre-Challenger data that shows that at 53-degree Fahrenheit the probability of joint failure is 0.8, and at freezing temperature, which was experienced during launch time, probability of failure is almost 1.0, thus in essence catastrophic.

Vaughan (1996) in her extensive and comprehensive analysis of the Challenger accident points out deficiencies within NASA's culture and Morton Thiokol, which prevented any of the working engineers from developing an awareness that there was a performance trend in which temperature might be important. She writes that the company failed to provide a continuity in engineering staffing to analyze post flight observational data "across all launches where temperature could be personally experienced." She writes that the events just prior to the Challenger launch showed that "professional accountability took a back seat" to "bureaucratic and political accountability." Vaughan's book points to a "normalization of deviance" and "group think" that had developed within NASA and Thiokol, which contributed to the accident. Namely, she argued that "it is possible that the process of deviance normalization …may play a role in facilitating rule violations and misconduct when they occur in the organization." Such organizational cultures and group thinking enable taking unnecessary risks in decision making processes. Vaughan (1996) points out the existence of a link between culture and individual choice, reported in scientific literature, where the culture factor is "a cause of organizational deviance and misconduct." She clearly puts "conformity" as the ultimate culprit of the Challenger Shuttle catastrophe. Vaughan (1996), based on that finding, strongly suggests incorporation of external objective evaluator in a decision making.

The Challenger accident happened despite, what Vaughan (1996) notes, NASA's ability to "create a decision-making structure absolutely suited to preventing the normalization of deviance." The NASA's decision making system did incorporate decentralized mode if needed but it did not prevent the accident.

Interestingly, Vaughan (1996) observed that many managers involved in the Challenger accident were former engineers and concludes that the new role gave them a new perspective on risk. Thus, changing the scope and character of job responsibilities can definitely affect risk tolerance.

Ann Skeet (2019) in her recent article on Boeing Max 737 plane accident points out "unhealthy corporate culture" as a main culprit and consequently proposes to increase the importance of the organizational culture and its management within organization. To signify the importance of a culture management as an ethics component she suggests to assign a "hard skill" attribute to it. Phil Hughes (2019) points out the age of the plane control design and plausible different internal views on the design technical aspects and consequently a different take on the airplane safety by Boeing engineering staff and management.

In a recently published work, Peacock (2014) investigated the normalization of deviance (NoD), the cause of the Challenger disaster proposed by Vaughan (1996) and how prevalent it is still in organizations based on empirical data. In particular he looked at the conflict between engineers and management from the perspective of psychological and sociological factors affecting decision making. He claimed that "engineers and managers could be argued to hold contrasting stakeholder perspectives linked to their respective professional obligations and exposure to risk." While engineers are considered "generally politically naïve" and averse to risk, management, due its obligation to stakeholders, has a much higher tolerance for risk. This anecdotal evidence was supported by Peacock's findings (Peacock, 2014). Namely, to Peacock's survey question: "If we followed their averse attitude to risk, no project would even go ahead, nor would we get anything done!" while only 24% of surveyed engineers answered "agree strongly" or "agree," 53% of surveyed managers answered positively. Peacock's chi-square test for association between role and risk aversion show a "statistically significant association" confirming that "engineers were more risk averse than managers." Peacock (2014) was looking at reasoning and the decision making process using the two perspectives utilized in the area of rationality of human cognition and decision making:

meliorism (in this approach human reasoning and decision making could be improved) and Panglossian method (human reasoning, competence and performance is almost always normatively correct) (Book Review, 2014.)

For a summary of findings of the discussed topics see Table 1 which differentiates engineers from managers with respect to various attributes related to ethical code paradigms, decision making, risk tolerance, rational reasoning, and value system.

Table 1. Engineers vs Managers: Roles, Rationality, Judgement Attributes (based on Peacock, 2014; Vaughan, 1996; Pinkus et al., 1997)

Attribute	Engineers	Managers
Codes of Practice	Professional ethics	More dichotomy especially when inequitable distribution of responsibility (IDR) is involved
Competency	Professional / technical	Managerial w/appreciation of technical data if needed
Responsibility	Professional / technical	Organizational / admin
Cicero's Creed II*	High adherence	Considered to various degree
Model of judgement	Tend to be more meliorist	Tend to be more Panglossian
Susceptibility to group think within organizational culture	Low	Medium to high
Risk tolerance	Low	High
Rationality /value system	Tend to be epistemic rationality / deliberation of facts	Instrumental rationality / focus on most effective to achieve a specific end
Decision process	Design: subjective / intuitive based on experience	Tend to be meliorist
Accountability nature	Professional	Bureaucratic / political
Conformity to group rules and norms	High if agrees w/professional ethics	High adherence to organizational culture / rules & norms
Inequitable distribution of responsibility (IDR)	Less applicable / more reluctant to apply	More prone to apply

*Note: Cicero's Creed II: The engineer should be cognizant of, sensitive to, and strive to avoid the potential for harm and opt for doing good (Pinkus et al., 1997)

Peacock's (2014) analysis points to managers' instrumental rationality as a "smoking gun," the most likely culprit for normalization of deviance experienced by organizations. He found that large firms show a higher propensity to develop normalization of deviance, which promotes a higher probability of failure. Peacock (2014) reports on cases where the normalization of deviance (NoD) was induced either directly ("demands imposed by management leading to compromising normative procedures") or indirectly ("erroneous procedures").

In summary, Table 2 contrasts a hypothetical small company with a large company with respect to various attributes of decision making discussed above, based partially on the findings of Peacock (2014).

Table 2. Small vs Large Firms: Decision Making Attributes

Attribute	Small company	Large Company
culture	Varies / more prone to admin subjectivity / more organic	Tend to be strong formal culture (which promotes melioristic decision making) / more mechanistic
operation	More experimental type	Tend to be operational
Judgement/decision	Varies: meliorist to Panglossian	Tend to be more meliorist
NoD	Lower tendency to occur	Higher tendency to occur

Project management process and its affecting factors have been subject of analysis by many disciplines including social sciences, psychology, and business among others that tried to understand formation of trust, effectiveness, and inner dependencies in decision making process (McAllister, 1995; Gapinski, 2017a,b, 2018; Robert, 2015). The findings point again to the importance of organizational culture among other factors and its impact on an effectiveness of the project management processes.

Deming (1987), in his approach to quality, stressed importance of systemic changes within any organization to improve quality, not merely watching production floors and seek attainment of quotas or meeting the production specifications. Had his philosophy of continuous improvement and open communication been followed, the rational decision making might have prevailed in overriding political conformity during the events leading to Challenger disaster.

6. Risk Assessment and Cumulative Failure Risk Factor

The literature on risks in a project, especially in the business discipline, provides risk assessment methodologies that incorporates probability assessment of events and their impacts in various formats. Larson and Gray (2014) provide a matrix that shows the impact scale of risk on project categories such as cost, time, scope, quality. Their risk severity matrix (RSM) graphically shows risk likelihood vs impact designed for specific risks. Similarly, Gido and Clements (2015) illustrate an example of risk assessment matrix for a specific web-based reporting project. To assess a risk of an event, the failure mode and effect analysis (FMEA) formula, that extends RSM method can be used (Larson and Gray, 2014). The multiplicative formula used for FMEA which considers severity of failure impact, probability of occurrence, and easiness of failure detection allows to assign a quantitative numerical value to a risk of an event to occur. Furthermore, to assess the project risks there are statistical techniques and tools available to project management although they focus mainly on financial risks.

All these methods and forms mentioned above usually describe risk assessment for very specific projects and lack universality, lack an additive, accumulative effect of risks, and don't take into account the factors described in this article that affect failure risks more profoundly. The models are devoid of critical factors such as organizational culture, communication, competency among others that affect decision making and, ultimately, determine the success or failure of a project. Consequently, the next segment defines the Cumulative Failure Risk Factor (CFRF), which can assist organizations in overall risk of failure assessment and the prompting of remedial actions.

The above analysis of the contributing factors affecting the project outcome motivated the author to propose the following model for the Cumulative Failure Risk Factor (CFRF). A calculated value of the CFRF serves as an indicator of the risk of project failure with a value expressed in percent determined by values assigned to the following contributing factors: Risk Tolerance (management's view), Communication Channels (open/top-bottom/bottom-up), Competency/Technical (at all levels at organization), Groupthink/Conformity, Cost/Time Overrun.

Thus, using the following labels and abbreviations for risks associated with contributing factors and their corresponding coefficients:

1. α coefficient for Risk Tolerance (RT)
2. β coefficient for Communication Channels (CCH)
3. γ coefficient for Competency/Technical (CT)
4. δ coefficient for Groupthink/Conformity (GC)
5. ε coefficient for Cost/Time Overrun (CTO)

the Cumulative Failure Risk Factor (CFRF) can be assessed as (equation 1):

$$\text{CFRF} = \alpha + \beta + \gamma + \delta + \varepsilon. \qquad (1)$$

where: α, β, γ, δ, ε are risk factor coefficients to be determined by the local project circumstances.

The sum of all coefficients cannot exceed, naturally, 100 percent. If all contributing factors have the same importance each of the coefficient is assigned value not exceeding 20 percent. The sum of all coefficients in any scenario must not exceed 100 percent, i.e., $\alpha + \beta + \gamma + \delta + \varepsilon <= 100\%$. It is assumed that contributing factors as defined are independent and mutually exclusive. Graphically the CFRF is described on Figure 1. The simplicity of CFRF allows for relatively fast risk assessment and corrective actions to be taken at any stage of the project execution.

Each contributing factor coefficient, i.e., α, β, γ, δ, and ε, is to be evaluated and assigned value in percent based on perceived risks of each of the contributing factors of RT, CCH, CT, GC, and CTO, respectively. The assigned value represents an actual risk towards project failure posed by a particular factor within the organization in the eyes of an evaluator for example, a high risk tolerance assumed by management poses a high probability for overall project failure. Consequently, the α coefficient for the RT factor should reflect it and be of much higher value than other factors, exceeding 20 percent. A poor communication culture within the organization, not promoting open communication in either top-bottom or bottom up setting would assign a higher value to the β coefficient for CCH factor than the others. A good communication climate, on other hand, would pose a much lower risk of a project failure by detecting shortcomings early on and the value assigned to the β coefficient for CCH factor should be low, close to 0%. The factor of competency within the organization encompasses the technical, organizational, and managerial competency. If competency including technical aspect is high the γ coefficient for CT factor should be low in value reflecting low risk of failure coming from that factor. The GC factor, representing groupthink and conformity reflects a risk of failure posed by these shortcomings. So, if a groupthink/conformity climate prevails within an organization, the value assigned to the δ coefficient for GC factor should be higher than others. In the case of costs and time overruns in a project, the ε coefficient for CTO factor should carry a higher value than others. If risks are distributed uniformly without a dominant factor, each individual coefficient value should not exceed 20 percent of contribution, otherwise a dominant risk factor value may exceed significantly all others.

Furthermore, based on fact that a risk tolerance may differ dramatically among various entities within an organization, it is advisable to assess the CFRF separately by different departments (management, engineering, etc.) prior to deriving final conclusions. The process of comparing and obtaining the consensus regarding overall project failure risk within an organization may bring additional useful insights and prompting corrective actions.

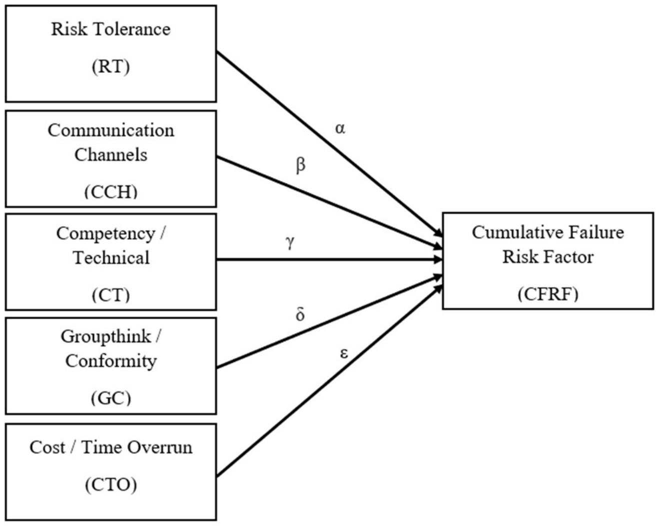


Figure 1. Cumulative Failure Risk Factor and its contributing factors

The CFRF through individual contributions of risk factors is tailored to individual circumstances and needs of the involved project. The ability to individually changing the importance of CFRF affecting factors gives organizations and project management's entities the flexibility to address the known or unknown but suspected project shortcomings or weaknesses.

While most of the evaluation of risk factors are usually performed internally within an organization, following Vaughan's recommendation, it is suggested to involve external viewpoints. As Vaughan (1996) writes, the external evaluators can deliver "outside the box" observations in determining the risk factors and consequently may assist in finding the numerical values assigned to the CFRF contributing factors. Thus, it is suggested to include the external evaluators to assess the risks and establishing CFRF value. This will provide an external and more objective reference for risk assessment within organization.

Adopting the occurrence probability scale after Engert and Lansdowne (1999) as:

0 – 10% very unlikely to occur
11 – 40% unlikely to occur
41 – 60% may occur about half of the time
61 – 90% likely to occur
91 – 100% very likely to occur,

it is suggested to assign threshold values for CFRF assessment as follows: a CFRF value between 0 and 40% as a low risk, between 41 and 50% represents a warning, between 51 and 60% a strong warning, any value in the range of 61% and 90% may indicate a high risk of failure occurrence, and a value above 91% represents a very high probability of a catastrophic failure. One may develop a more detailed scale depending on local needs such as in Garvey (2001). The scale and specific value ranges are left to local settings and interpretation.

In order to check the performance of the proposed risk of failure assessment model the Challenger case was considered in two assessment simulation scenarios: one performed by "insiders" and one performed with inclusion of external evaluators based on available literature and reports. The author obtained the following results: $CFRF_1$ = in the range of 50 - 65% for first scenario, where the project technical aspect and conformity issues were not internally fully recognized, and $CFRF_2$ = in the range of 90 - 100% for second scenario, where technical aspect of high probability of booster joint failure and high conformity in organizational culture were realized. In both cases it is worth of mentioning that the higher risk tolerance displayed by management comparing to engineering staff that was caused by political pressures affected CFRF values as well. The second value of $CFRF_2$ points to a very high probability of a failure, which agrees with result of Pinkus et al. (1997) probability estimation model. The proposed model to assess a risk of failure is currently subject of further testing.

7. Conclusions

The article analyzes engineering project failures and disasters. The article provides examples from antiquity to modern times covering civil engineering, construction, electrical and information technology sectors. It identifies and reviews project failure causes and provides their classification. The author analyzes engineering and management interdependencies and conflicting demands placed upon them, which may result in a catastrophic failure. The Challenger Shuttle disaster is analyzed from the perspective of conflicting demands and expectations faced by management and engineering. Shortcomings in communication, conformity, decision making, a culture of arrogance are analyzed as failure culprits. The cumulative failure risk factor (CFRF), which considers risk tolerance, communication, competence, groupthink, and overruns aspects, is proposed, subject to individual settings, to assess a project's failure risk.

8. References

Altay, N., & Green III, W. (2006). OR/MS research in disaster operations management. European Journal of Operational Research. 175. Pp. 475-493.

Beauvais Cathedral (1284). https://en.wikipedia.org/wiki/Beauvais_Cathedral

Bhopal (1984). https://en.wikipedia.org/wiki/Bhopal_disaster (accessed Oct. 21, 2018)

Blackout (2003). https://en.wikipedia.org/wiki/Northeast_blackout_of_2003. (Accessed Oct. 27, 2018)

Book Review (2014). Book review of Stanovich, K. E. Rationality and the Reflective Mind. New York. Oxford University Press. (2011) https://www.tandfonline.com/doi/abs/10.1080/09515089.2013.766166?mobileUi=0&journalCode=cphp20 (accessed Nov.1, 2018)

Chapman, C. and Ward, S. (2004). Project Risk Management Processes Techniques and Insights. 2nd ed. John Wiley & Sons.

Chernobyl (1986). https://en.wikipedia.org/wiki/Chernobyl (accessed Oct. 20, 2018)

Crouhy, M., Galai, D., and Mark, R. (2005). Essentials of Risk Management. McGraw-Hill.

Deepwater Horizon (2010). https://www.britannica.com/event/Deepwater-Horizon-oil-spill-of-2010 (accessed Oct. 20, 2018)

Deming (1984). https://www.youtube.com/watch?v=0yGhR1ybmN8 (accessed Oct. 10, 2018)

Discenza, R., & Forman, J.B. (2007). Seven causes of project failure: how to recognize them and how to initiate project recovery. Paper presented at PMI Global Congress 2007 – North America, Atlanta, GA. Newton Square, PA: Project Management Institute.

Egeland, B. (2012). Project Recovery Strategies http://pmtips.net/project-recovery-strategies/

Electrical Disasters (2018). http://thinkcopper.org/sustainability/4-steps-to-avoiding-electrical-disasters/ (accessed Oct. 27, 2018)

Engert, P.E., Lansdowne, Z.F. (1999). Risk Matrix User's Guide. Version 2.2. Mitre Document MP99B0000029. Nov. The Mitre Corporation.

Gapinski, A. (2019). On Factors Affecting Trust in Project Management. Research in Logistics and Production. Int. Journal. Vol. 9. No. 1. January Issue. ISSN 2083-4942 (Print) ISSN 2083-4950 (Online) DOI: 10.21008/j.2083-4950.2019.9.1.4. Pp. 39-47.

Gapinski, A. (2017a). Trust in Project Management. Journal of Management and Engineering Integration. Vol. 10, Issue 2. Winter Issue. Pp. 1-11.

Gapinski, A. (2017b). Trust in Project Management: Analysis of the Proposed Model. Journal of Management and Engineering Integration. Vol. 10, Issue 2. Winter Issue. Pp. 12-19.

Garvey, P.R. (2001). Implementing a Risk Management Process for a Large Scale Information System Upgrade – A Case Study. INCOSE Insight. May.

Gido, J., and Clements, J.P. (2015). Successful project Management. 6th ed. Cengage Learning. Stamford, CT.

Hall, P. (1980). Great Planning Disasters. London: George Weidenfeld and Nicolson Ltd.

Heldman, K. (2005). Risk Management. John Wiley & Sons.

Hornjak, B. (2001). The Project Surgeon. Newton Square, PA: Project Management Institute.

Hughes, P. (2019). Did Boeing's managers override its engineers? http://www.ft.com/content/b8d9c916-5c82-11e9-9dde-7aedca0a0a081a (accessed April 17, 2019).

Kliem, R. (2011). Managing projects in trouble: Achieving turnaround and success. Boca Raton, FL. Auerbach publications.

Koetz, M. (2018). Donora Smog: How Nuclear Energy Cleared The Air. Tribune-Review. A7. Oct. 21.

McAllister, D.J. (1995). Affect-and Cognition-Based trust as Foundations for Interpersonal Cooperation in Organizations. Academy of Management Journal. Vol 38, No. 1. Pp.24-59.

McDonald, A.J., with Hansen, J.R. (2009). Truth, Lies, and O-Rings. Inside the Space Shuttle Challenger Disaster. University Press of Florida.

Mocranz, P.D., and Taylor, R.K. (2000). Engineering Process Failure – Hyatt Walkway collapse. ASCE Journal of Performance of Constructed Facilities, 14(2). Pp. 46-50.

Morris, P., & Hough, G. (1987). The Anatomy of Major Projects. Chichester: John Wiley & Sons.

Larson, E.W., Gray, C.F. (2014). Project Management. The Managerial Process. 6th ed. McGraw-Hill Education. New York, NY.

Peackock, S. (2014). Managers and Engineers. A Multi-perspective Analysis of Conflict. University of Leeds. UK. MBA Master Thesis.

Pinkus, R.L.B., Shuman, L.J., Hummon, N.P., and Wolfe, H. (1997). Engineering Ethics. Balancing Cost, Schedule, and Risk – Lessons Learned from the Space Shuttle. Cambridge University Press.

Quecreek Mine (2013). https://en.wikipedia.org/wiki/Quecreek_Mine_rescue (accessed Oct. 15, 2018).

Rae, T., Eden, C. (2002). On Project Success and Failure in Major Engineering Projects. Submitted to the European Academy of Management (EURAM 2002).

Rialto Bridge (1444). https://en.wikipedia.org/wiki/Rialto_Bridge

Robert, D. (2015). The Effects of Trust on the Effectiveness of Project Risk Management for Engineering and Construction Projects. Master Thesis. Delft University of Technology.

Rogers Commission (1986). https://spaceflight.nasa.gov/outreach/SignificantIncidents/assets/rogers_commission_report.pdf

Rogers, S. (2014). Ancient Engineering Fail: Historic Structural Disasters. https://weburbanist.com/2014/04/16/ancient-engineering-fail-12-historic-structural-disasters/

Rosencrance, L. (2007). Survey: Poor Communication causes most IT project failures. Computerworld. https://www.computerworld.com/article/2543770/it-management/survey--poor-communication-causes-most-it-project-failures.html

Skeet, A. (2019). Opinion: Here's what we can learn from Boeing's unhealthy corporate culture. https://www.marketwatch.com/story/heres-what-we-can-learn-from-boeings-unhealthy-corporate-culture-2019-04-17?siteid=yhoof2&yptr=yahoo (accessed April 17, 2019).

Storm, P. & Savelsbergh, C. (2014). Project recovery management. Paper presented at Project Management Institute Research and Education Conference, Phoenix, AZ. Newtown Square, PA: Project Management Institute.

Tacoma (1940). https://en.wikipedia.org/wiki/Tacoma_Narrows_Bridge_(1940) (accessed Oct. 21, 2018).

Thean, L.P., Vijiaratnam, A., Lee, S.L., and Bengt, B.B. (1987). Report of the Inquiry into the Collapse of Hotel New World. Singapore National Printers. Singapore.

Three Mile Island (1979). https://en.wikipedia.org/wiki/Three_Mile_Island_accident (accessed Oct. 21, 2018)

Thompkin, P.K. (2005). Apollo, Challenger, Columbia: The Decline of the Space Program. Roxbury (now Oxford University Press).

Transformers (2018). https://www.smithsonianmag.com/smart-news/what-makes-transformers-explode-101228674/

Turner, J., & Müller, R. (2006). Choosing Appropriate Project Managers. Newton Square, PA: Project Management Institute.

Vaughan, D. (1996). The Challenger Launch Decision. Risky Technology, Culture, and Deviance at NASA. The University of Chicago Press. Chicago and London.

Weick, K., & Sutcliffe, K. (2001). Managing the Unexpected. San Francisco: Jossey-Bass.

Willow Island (1940). https://en.wikipedia.org/wiki/Willow_Island_disaster (accessed Oct. 18, 2018)

Workfront (2018). Project Failure: 10 Famous Failures and 5 Ways to Spot Them Before They Happen. http://www.workfront.com/blog/project-failure-10-famous-failures-and-5-ways-to-spot-them-before-they-happen (retrieved September, 15, 2018)

Lean in Education: Mistake-Proofing Methods Used by Teachers at a Magnet High School

Robert S. Keyser[1]
[1]*Kennesaw State University*
rkeyser@kennesaw.edu

Abstract

High school teachers, like most people, make mistakes on a regular basis, but they have also developed mistake-proofing methods to negate the impact of these errors. There is currently no published research in the field of identifying common mistakes and employing mistake-proofing techniques among teachers in a high school environment. A qualitative, descriptive study describes the most common mistakes made by teachers at a magnet high school in the metro-Atlanta area as well as the mistake-proofing methods used to mitigate these mistakes. An online survey was distributed to all faculty at the high school and 30 responses were received within the four-week deadline. The most common error reported related to the grading of assignments. Teachers also struggled with personal time management, as well as classroom time management. Results indicate six common types of mistakes and their respective mistake-proofing methods that are discussed in this paper.

1. Introduction

Mistake-proofing (or error-proofing) is a Lean concept dating back to the 1600s, but was popularized by Shigeo Shingo and Taiichi Ohno during the creation of the Toyota Production System (TPS) (Liker, 2004). Since Toyota's adoption of mistake-proofing, many corporate businesses have also adopted mistake-proofing (Schmidt, 2013). Now, medical practices have begun to experiment with mistake-proofing methods (Grout, 2013).

What constitutes a mistake, or error? In manufacturing, some consider a mistake to be anything that creates waste (Liker, 2004). In quality control, a mistake is a defect or fault in the product (Goldratt, 2007). For purposes of this research, the researchers have adopted Goldratt's definition and conveyed to the teachers that a mistake (or error) is a defect or fault in the product, the intent being that if a teacher's mistake takes away from a student's quality of education, it should be corrected (Kuyini, 2007).

There are numerous ways to implement mistake-proofing methods. Typically, mistake-proofing efforts have focused on repeatable tasks, the flow of both materials and information, and the production flow, such as those found in manufacturing (Nicolay, 2012). Moreover, there are three categories of mistake-proofing: (1) prevention in the environment; (2) error detection to prevention; and (3) prevention of the influence (Grout, 2007). Prevention of mistakes in the environment employs a device inside a system to send visual displays or notices about an error (Schmidt, 2013). An example would be placing tiles on the floor that signal certain restrictions in the area (Grout, 2007). Workers tend to use error detection to prevention when management encourages workers to find errors and then develop unique solutions to avoid or correct the error (Liker, 2004). An example of prevention of the influence of an error is when a worker detects an error and then fixes the error before the product moves further along the production queue (Goldratt, 2014).

The impact of mistake-proofing is inherently different for each organization. At Toyota, for example, implementing mistake-proofing into a former General Motors plant increased job satisfaction (Liker, 2004). Mistake-proofing minimized the quantity of mistakes made, thereby reducing stress levels, in the logistics industry (Schmidt, 2013). In the healthcare industry, changing a medical kit increased the price of the medical kit, but halved the number of infections which saved an organization approximately three million dollars (Grout, 2007). In the food service industry, mistake-proofing methods increased worker job satisfaction and morale (Keyser, 2017). In short, mistake-proofing improves the quality of products and reduces waste (Liker, 2004; Manivannan, 2006). Manivannan further suggests that organizations should be focused on continuous development of techniques and systems to avoid future errors in order for mistake-proofing to have a lasting impact.

Teachers, like most people, routinely make mistakes (or errors) despite advances in error-proofing mechanisms such as Spell Check and proofreading of manuscripts. Mistakes still tend to slip through the cracks and, if mistakes occur often enough, this may lead to an increase in stress levels and job dissatisfaction for the teacher. Examples of increased student stress levels resulting from teacher mistakes include: 1) if the teacher inadvertently failed to cover the material on an assessment, thereby creating angst for the student while taking the assessment, such as an exam; and 2) if incorrect grades were entered in the teacher’s gradebook for their respective assignments, thereby perplexing students who thought they had performed better than their grades reflect on their periodic grade reports. Further, increased stress levels for both the teacher and student could lead to severe health-related issues. If teachers are sufficiently dissatisfied with their jobs, this may lead to their seeking other employment, thus, having a deleterious effect on teacher retention rates. However, like many people in the industrial world, teachers have developed mistake-proofing methods to make their jobs easier. Despite the growing number of teachers using mistake-proofing techniques, there exists no current published literature of the mistake-proofing methods that teachers use. The purpose of this study is to describe the mistakes that teachers make and their resultant mistake-proofing methods developed at an Atlanta-area Magnet High School that specializes in drawing academically motivated students to their Academy of Mathematics, Science, and Technology program. The scope of this study is to learn what types of mistakes are common to teachers and the mistake-proofing methods that teachers develop to address these mistakes. This provides valuable insight into how mistake-proofing is employed in higher education.

2. Literature review

The successful applications of lean techniques in the manufacturing sector are widely published. Although Lean sometimes fails in organizations (Keyser et al., 2016), most industrial practitioners will attest to the many benefits afforded by successful Lean implementation. Whereas published literature has demonstrated Lean applications in the healthcare, logistics, marketing and sales industries, Lean can also be successfully applied in other industries such as the food service industry and education.

De Steur et al. (2016) conducted a systematic review to show the potential of value stream mapping, Just-in-Time, and 5s to identify and reduce food losses and wastes in the supply chain. An in-depth bibliographic study by dos Reis Leite (2015) reveals that applying the Lean philosophy incorporates a mix of tools and practices to a given situation in the services sector. Malihe et al. (2014) presents a collection of Lean tools and techniques, such as demand collaboration, continuous improvement, and inventory management, that are vital for decreasing costs and wastes on Halal food supply chains. Farhad & Alireza (2009) use six sigma methodology to reduce delivery lead time in a food distribution network for small- to medium-sized enterprises (SME). The use of value chain analysis in the supply chain highlights waste elimination opportunities at both the producer and processor levels in the UK beef foodservice sector (Francis et al., 2008). (Keyser et al., 2017) discusses how Lean was applied in the food service industry citing three restaurant examples in Tennessee using tools and techniques such as mistake-proofing, visual management, 6s, kaizen, and setup time reduction,

Although published research concerning the application of Lean in higher education institutions exists, current published research is bereft of mistake-proofing common errors in higher education institutions. Sunder and Antony (2018) suggest a six-stage conceptual framework for applying lean/six sigma (LSS) in higher education institutions (HEIs). The six stages center around involvement from top management in ascertaining whether the organization is ready and aligned to carry out the organizational vision for quality excellence and, if so, a deployment strategy must then be developed. Stakeholders must be educated and teams formed to carry out the deployment strategy. Identifying and implementing LSS projects, then executing the conclusions is the final critical stage. They do recognize that this is a conceptual framework, limited by the fact that it had not been tested in a real-life context.

Narayanamurthy, et al. (2017) developed a framework that provides a structured procedure for implementing Lean thinking (LT) in educational institutions by using an action research methodology to identify wastes and propose various solutions in the initial stages of implementation.

Doman (2011) allowed undergraduate students to apply and reinforce what they learned about Lean in their coursework towards improving administrative processes in higher education. LeMahieu et al. (2017) engaged an organization in continuous problem-solving, learning, and quality improvement with Plan-Do-Check-Act (PDCA) cycles by improving technology support and services for instructional purposes in a school district system. Maguad (2007) used Lean techniques to reduce waste and improve operational efficiency. Stewart and Grout (2001) believes that an understanding of the cognitive mechanisms underlying human error will aid in the prevention of mistakes by adjusting the work environment to provide cues that prevent a particular cognitive response. Teachers were more likely to remain in their jobs if there were professional development opportunities, student interaction, more available resources, and being able to make mistakes without fear of repercussion (Gaikhorst et al., 2014).

3. Design Methodology

The research design for this study is a qualitative descriptive design, which allows for exploratory research into the types and causes of mistakes by teachers, and mistake-proofing methods developed by the teachers. The research method employed was an online survey consisting of eight demographic questions and two open-ended questions distributed to all faculty working at the magnet high school. A convenience sample of n = 30 responses were received. A convenience sample is such that data is obtained based on availability rather than as representative of the population (Leedy, 2001) – simply put, 30 responses were received and all 30 responses were used for analysis in this study.

The survey was modified from a study regarding the errors made by election officials (Ansolabehere & Shaw, 2016). The two open-ended questions were modified to pertain to the teaching profession. All participants were asked to identify three mistakes they routinely made and the mistake-proofing methods they used for each mistake.

With assistance from the high school research and internship director, all faculty were provided with a link to the survey via email. The survey was available for four weeks, whereupon the categorical data was then entered into a spreadsheet for analysis.

4. Analysis

The purpose of this study is to describe the mistakes that teachers make and their resultant mistake-proofing methods developed at an Atlanta-area Magnet High School. From the n = 30 responses in the survey, a total of 82 matched sets – that is, 82 mistakes accompanied by their respective mistake-proofing methods were used in the analysis. One odd set, meaning one mistake was stated but it lacked a mistake-proofing method, was eliminated from further consideration. The results are shown in the order of the questions in the survey.

Question 1: All participants agreed to the terms and conditions listed in the Participant Consent Form. The terms and conditions pertained to the fact that participation was strictly voluntary and anonymous and that agreement allowed consent for the researchers to use this data in the study.

Question 2: This question asked for three common mistakes made by the participant. The 82 mistakes cited by participants were categorized into six common mistake types: 1. Assignment, 2. Attendance, 3. Grading, 4. Management, 5. Misjudgment, and 6. Presentation. The category with the highest frequency of mistakes was ***Grading*** with 22 responses out of the total of 82 responses. Examples of grading mistakes include unfair grading of open-ended questions, mis-graded multiple-choice problems, and grade entry mistakes. The second largest category was ***Management*** with 20 responses. Management referred to both time management and management of the classroom. The third largest category of mistakes was ***Presentation*** with 15 responses. Presentation mistakes consisted of mistakes by teachers during class lectures or information that was presented in one class but not in other similar classes. The fourth largest category of mistakes was ***Assignment*** with 12 responses. Examples include typographical mistakes in the assignments, poorly designed assignments, and nebulous language causing student confusion in understanding the assignment. The fifth largest category of mistakes was ***Misjudgment*** with 9 responses. Misjudgment mistakes are related to teachers either overestimating or underestimating the students' knowledge, motivation, or their access to resources to do the assignments. Finally, the sixth, and smallest, category of mistakes was ***Attendance*** with 4 responses. In all four cases, this mistake was attributed to simply forgetting to take attendance.

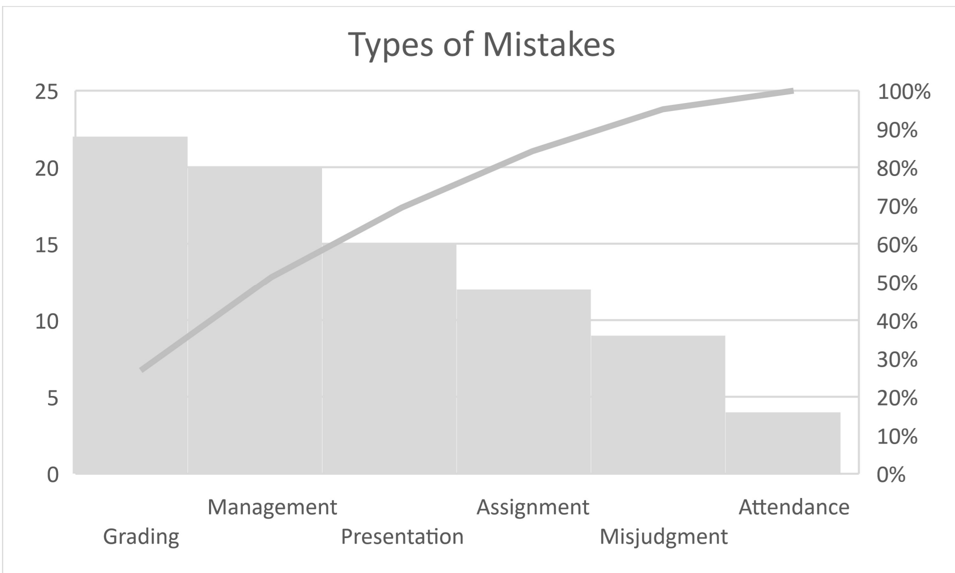


Figure 1. Pareto Chart of Types of Mistakes

Question 3: This question asked for three mistake-proofing methods – one for each mistake listed by the participant to indicate current mistake-proofing methods that teachers use. The 82 mistake-proofing methods cited by participants were categorized by highest frequency of the six common mistake types from Question 2: 1. Grading, 2. Management, 3. Presentation, 4. Assignment, 5. Misjudgment, and 6. Attendance.

Grading. Grading was decomposed into eight types of mistake-proofing methods: 1. Analysis, 2. Combine grades, 3. External Tools (i.e., ruler), 4. Pay attention, 5. Review grades, 6. Review answer key, 7. Grading rubric, and 8. Student check grades (for verification of correct grade). Of the 22 responses, Review grades and Student check grades for verification were tied with 5 responses each. These results were followed by using an External Tool such as a ruler to aid in following student grades by row down a spreadsheet and Review Answer Key to verify correct answers before grading with 3 responses each. To address the unfair grading of open-ended questions, there were two responses citing the creation of a grading rubric, rather than using subjective grading, as a remedy for this problem. Paying closer attention to grade entry (2 responses), Combining grades to make up for a missing grade (1 response), and further Analysis of inaccurately graded multiple choice questions (1 response) rounded out the remaining responses.

Table 1. Grading

Mistake-proofing methods	Count
Analysis	1
Combine grades	1
Use external tool (i.e., a ruler)	3
Pay attention	2
Review grades	5
Review answer key	3
Create a grading rubric	2
Student check own grades	5
Total	22

Management. Management was decomposed into four types of mistake-proofing methods: Constant adjustment, Effective use of planning time, Elimination of distractions, and Make it work. Constant adjustment led the Management category with 7 out of 20 total responses. Constant adjustment refers to constantly adjusting classroom management and poor plans for the day. Planning had 10 responses. The mistake-proofing method cited most frequently was to make time to plan ahead rather than at the last minute. Other mistake-proofing methods include making it work even with a lack of planning (2 responses) and eliminate distractions while planning (1 response).

Table 2. Management

Mistake-proofing methods	Count
Constant adjustment	7
Planning	10
Eliminate distractions	1
Make it work	2
Total	20

Presentation. The 14 responses for Presentation were decomposed into four types of mistake-proofing methods: Planning the lecture, Re-teach the topic, Creating an outline, and Reviewing the lecture. Six teachers used an outline to avoid giving students incorrect information and to ensure that all classes received the same information on the topic. There were five responses for reviewing the presentation to catch and correct any mistakes before giving the presentation to the class. Specifically, teachers were looking for typographical errors and checking to see that all necessary lecture details were covered in the presentation. Two responses indicated planning for specific information to be discussed in class that day and one response indicated re-teaching the lecture to students who did not comprehend the initial lecture on a topic.

Table 3. Presentation

Mistake-proofing methods	Count
Planning specific info	2
Re-teach topic	1
Outline	6
Review lecture	5
Total	14

Assignment. Assignments was decomposed into three types of mistake-proofing methods: Planning, Revision, and Creating the assignment correctly. Seven of the 12 total mistake-proofing responses referenced revising assignments after finding typographical errors and poorly worded instructions in the assignment. Planning followed with 4 responses. Planning referred to reducing both the number of mistakes made and the length of the assignment. One respondent cited creating the assignment correctly the first time to alleviate any problems associated with a given assignment.

Table 4. Assignment

Mistake-proofing methods	Count
Planning to reduce mistakes	4
Revise assignment	7
Create it right	1
Total	12

Misjudgment. Misjudgment was decomposed into three types of mistake-proofing methods: Revise assignment, Provide extra materials and assignments, and Reviewing the lecture material with students. Of a total of 10 responses, there were 4 responses for Assignment. Here, teachers who cited Revise assignments were trying to prevent both overestimating students' knowledge and underestimating students' self-motivation. Providing extra materials and assignments received 4 responses. This was designed to provide extra practice students needed to master a topic. The final 2 responses pertained to teachers who reviewed course materials with students prior to a test in order to ensure that students were adequately prepared to take the test.

Table 5. Misjudgment

Mistake-proofing methods	Count
Revise assignment	4
Provide extra materials/assign.	4
Review lecture material	2
Total	10

Attendance. Attendance was decomposed into three types of mistake-proofing methods: Lesson, Attention, and Students. Of the 4 total responses, adding time to the lesson (keeping students busy) to allow time to take attendance received 2 responses. Paying close attention to taking attendance and using students to confirm attendance (1 response each) were employed to overcome the mistake of inaccurately recording attendance.

Table 6. Attendance

Mistake-proofing methods	Count
Add time to lesson	2
Pay attention	1
Using students to confirm	1
Total	4

The following categories (in bold face type) below round out Questions 4-8 in the survey.
Demographics. Demographics for the participants in this study for Gender, Age, Department, and Experience and discussed below.
Gender. Of the *n* = 30 participants in this study, 20 were female and 10 were male.
Age Range. Four categories for age range and their respective counts are shown in Table 7.

Table 7. Age range of participants

Age Range	Count
25 to 34	2
35 to 44	11
45 to 54	15
55 to 64	2

Department. Eight departments at the magnet high school and the number of participants in each department are shown in Table 8.

Table 8. Departments of participants

Departments	Count
Career Tech	3
Fine Arts	2
Language Arts	4
Math	8
Physical Education	3
Science	5
Social Studies	4
Special Education	1

Experience. Five categories of number of years of experience and the number of participants are shown in Table 9.

Table 9. Years of experience of participants

Years of Experience Range	Count
0 to 5	4
6 to 10	2
11 to 15	7
16 to 20	6
Over 20	11

5. Research limitations

This study has the following research limitations: (1) Participant's understanding and awareness of a mistake. If a teacher is unable to recognize a mistake that he/she made, then it would be difficult for him/her to report it accurately. (2) Despite the survey being anonymous, a teacher may still be unwilling to admit a mistake. (3) The study cannot be generalized due to its small sample size of one magnet high school in the metro-Atlanta area with 30 responses. (4) Participation in the online survey was voluntary and anonymous. Therefore, the convenience sample size of n = 30 may not be considered representative of the entire faculty population. (5) If a teacher was unable to fully comprehend the definition of a mistake (or error) in terms of this study, then his/her response may not be as accurate as compared to the responses from other participants.

6. Conclusions and Recommendations

Thirty responses were received yielding 82 reported mistakes and their corresponding mistake-proofing methods. The most prevalent type of mistake reported was Grading. Examples of grading mistakes include unfair grading of open-ended questions, incorrect or missing grade entry into a gradebook, and mis-graded multiple-choice problems. Mistake-proofing methods included reviewing grades entered, using an external tool such as a ruler to help the teacher as he/she reads down a spreadsheet, reviewing the answer key for correctness prior to grading an assignment, and creating a grading rubric to fairly assess all student responses to open-ended questions.

The second most common type of mistake reported was Management – that is, time management and management of the classroom during a lecture period. There were instances when teachers felt there was too much data to analyze in the time given and also poor planning that contributed to the ineffective use of classroom time. Mistake-proofing methods included making time in advance to properly plan the lecture before presenting it followed by constantly making adjustments during the lecture period to cover requisite material despite a poorly planned lecture.

The third most common type of mistake reported was Presentation. Presentation mistakes referred to mistakes that teachers made during a lecture or the fact that a teacher would present a certain body of material in one class but forget to present this same material in other similar classes. The top mistake-proofing method was creating an outline to avoid giving students incorrect information and to ensure that all classes received the same information on the topic followed by reviewing the presentation to catch and correct any mistakes before giving the presentation to the class.

The fourth most common type of mistake reported was Assignment. Assignment mistakes referred to typographical mistakes in the assignments, poorly designed assignments, and nebulous language causing student confusion in understanding the assignment. Mistake-proofing methods included revising the assignment and also planning to reduce the number of mistakes in an assignment to zero.

The fifth most common type of mistake reported was Misjudgment. Examples of Misjudgment mistakes include teachers either overestimating or underestimating the students' knowledge, motivation, or their access to resources to do the assignments. Mistake-proofing methods included revising assignments that were designed to prevent both overestimating students' knowledge and underestimating students' self-motivation and providing extra materials and extra assignments to give students more practice in order to master a topic.

Finally, the sixth most common type of mistake reported was Attendance. Attendance mistakes primarily occurred when teachers forgot to take attendance. Mistake-proofing methods included paying closer attention to taking attendance in class and also adding time to lectures in order to keep students busy to allow time for the teacher to take attendance.

Recommendations include conducting future studies of multiple high schools and to use random sampling, if the sample size is sufficient, to determine whether those teacher mistakes align with what was discovered in this seminal research study, thus, generalizing the results on a much broader scale. This type of research study could also be conducted at the college or university level.

7. References

Ansolabehere, S., & Shaw, D. (2016). Assessing (and fixing?) election day lines: Evidence from a survey of local election officials. Electoral Studies.

De Steur, H., Pearce, D., Wesana, J., Dora, M. K., & Gellynck, X. (2016). Applying value stream mapping to reduce food losses and wastes in supply chains: A systematic review. Waste Management, 58, 359-368

Doman, M. S. (2011). A new lean paradigm in higher education: a case study. Quality Assurance in Education, 19(3), 248-262.

Dos Reis Leite, H. (2015). Lean philosophy and its applications in the service industry: A review of current knowledge. Production, 25(3), 529-541.

Francis, M., Simons, D., & Bourlakis, M. (2008). Value chain analysis in the UK beef foodservice sector. Supply Chain Management, 13(1), 83-91.

Gaikhorst, L., Beishuizen, J. J., Korstjens, I. M., & Volman, M. L. (2014). Induction of beginning teachers in urban environments: An exploration of the support structure and culture for beginning teachers at primary schools needed to improve retention of primary school teachers. Teaching and Teacher Education, 42, 23-33.

Goldratt, E. M., & Cox, J. (2014). The goal: A process of ongoing improvement. Great Barrington, MA: North River Press.

Grout, J. (2007). Mistake-proofing the design of health care processes. Rockville, MD: Agency for Healthcare Research and Quality.

Keyser, R. S., Marella, V.K., & Clay, K. (2017). Lean restaurants: Improving the dining experience. Journal of Higher Education Theory and Practice, 17(7), 67-79.

Keyser, R. S., Sawhney, R. S., & Marella, L. (2016). A management framework for understanding change in a lean environment, Tékhne - Review of Applied Management Studies, 14(1), 41.

Kuyini, A. B. (2007). Ghana's education reform 2007: A realistic proposition or a crisis of vision? International Review of Education, 59(2), 157-176.

Leedy, P. (2001). Practical research: Planning and design. Upper Saddle River, NJ: Merrill Prentice Hall.

LeMahieu, P. G., Nordstrum, L. E., & Greco, P. (2017). Lean for education. Quality Assurance in Education: An International Perspective, 25(1), 74-90.

Liker, J. K. (2004). The Toyota way: 14 management principles from the world's greatest manufacturer. New York, NY: McGraw-Hill.

Maguad, B. A. (2007). Lean strategies for education: Overcoming the waste factor. Education, 128(2), 248-255.

Malihe, M., Ab-Rahman, M. N., Zain, C. R. C. M., & Jamsari, E. A. (2014). Increasing production and eliminating waste through lean tools and techniques for Halal food companies. Sustainability, 6(12), 9179-9204.

Manivannan, S. (2006). Error-proofing enhances quality. Manufacturing Engineering, 137(5), 99-104.

Nabhani, F., & Shokri, A. (2009). Reducing the delivery lead time in a food distribution SME through the implementation of six sigma methodology. Journal of Manufacturing Technology Management, 20(7), 957-974.

Narayanamurthy, G., Gurumurthy, A., & Chockalingam, R. (2017). Applying lean thinking in an educational institute – an action research. International Journal of Productivity & Management Performance Management, 66(5), 598-629.

Nicolay, C. R., Purkayasthat, S., Greenhalgh, A., Benn, J., Chaturvedi, S., Phillips, N., & Darzi, A. (2012). Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. The British Journal of Surgery, 99(3), 324-335.

Schmidt, S. (2013). Preventive methods in logistics poka-yoke and failure mode and effect analysis (FEMA). Acta Technica Corvininesis – Bulletin of Engineering, 6(1), 27-30.

Stewart, D. M., & Grout, J. R. (2001). The human side of mistake-proofing. Production & Operations Management, 10(4), 440-459.

Sunder,V. M., & Antony, J. (2018). A conceptual Lean Six Sigma framework for quality excellence in higher education institutions. International Journal of Quality & Reliability Management, 35(4), 857-874.

Review of Earned Value Management (EVM) Methodology, its Limitations, and Applicable Extensions

Anisulrahman Nizam[1]
Dr. Ahmad Elshannaway[1]
[1]*University of Central Florida*
anizam@knights.ucf.edu ; ahmad.elshannaway@ucf.edu

Abstract

Department of Defense (DoD) Instruction 5000.02 requires an Earned Value Management System (EVMS) compliant with ANSI/EIA-748 for all DoD cost or incentive contracts valued at or greater than $20M. Earned Value Management (EVM) integrates cost, schedule, and time to draw conclusions about current project status as well as make projections for future project status. Though EVM has been widely adopted on many projects, there are clear limitations indicated in the literature which ultimately inhibit the ability of EVM to become universally accepted as a best practice across all industries. In response, researchers have developed extensions such as Earned Schedule Management (ESM), Earned Duration Management (EDM), and Customer Earned Value (CEV). This paper addresses the evolution, limitations, and new extensions of EVM.

1. EVM Introduction

Earned Value Management (EVM) is a comprehensive performance measurement system (PMS) that integrates cost and schedule parameters into a single methodology to provide joint situational awareness for project managers and customers to assess project cost, schedule, and technical performance. EVM is an increasingly popular tool that organizations are utilizing to report and control project performance in an objective manner. EVM is mostly prevalent in the defense industry as it has been mandated by the United States Government for DoD contracts valued at or greater than $20M in accordance with ANSI/EIA-748. Utilized as a PMS, EVM helps drive organizational success (Upadhaya, 2014). Without an effective PMS, an organization lacks the ability to track, monitor, or take corrective actions as necessary. Prior to EVM, traditional PMS's had two separate and independent systems with one focusing on cost and the other focusing on schedule. Lacking integration of these two systems, a project manager could not truly understand the health of the project. This glaring weakness could not identify the reason a project was over or under spending since it did not cross reference time-based data. EVM brings together cost and schedule data by integrating them into one metric. By integrating cost and schedule data together, project managers and the contracting agency can monitor project health while providing a mechanism to forecast the final cost at completion of the project as well as when the project will be completed. This easily aligns the organization at both the strategic and operational levels (McAdam, 2014) by providing detailed day to day information while also providing the performance of the project overall. EVM does this by using a common monetized value of work for both cost and schedule. This is where the true power of EVM lies, allowing EVM to produce variance and performance indices to predict final project cost and schedule at completion.

2. Basic Definitions of EVM

There are three fundamental metrics of EV that are used to generate the performance indices for cost performance and schedule. They are Planned Value (PV), Earned Value (EV), and Actual Costs (AC). PV is defined as the budgeted cost of work scheduled at the measuring point (MP). EV is defined as the budgeted cost of work performed at MP. Finally, AC is defined as the actual cost of work performed at the MP. Once these three fundamental measures are obtained, several key metrics and ratios can be derived which indicate the health, performance, and outlook of the project. The Cost Variance (CV) is equal to the difference between EV and AC. This number indicates the extent of over or under run in terms of cost. The Schedule Variance (SV) is equal to the difference between EV and PV. This number indicates whether a project is running behind or ahead of schedule. Negative values of CV and SV indicate a lack of progress against the baseline plan. The Cost Performance Index (CPI) is equal to the EV divided by AC. This is a powerful index that indicates how much it costs to earn one dollar of budget (Wake, 2008). The Schedule Performance Index (SPI) is equal to EV divided by PV. This index indicates the extent to which the project is running ahead or behind schedule. If both SPI and CPI are equal to 1, the project is running exactly on schedule and budget. The To Complete Performance Index (TCPI) is a useful index that indicates how well a project must perform in terms of cost on the remaining work in order to complete on budget. The TCPI is equal to the difference in Budget at Completion and EV divided by the difference in Budget at Completion and AC. As a general rule of thumb, once a TCPI exceeds a value of 1.1, such a dramatic shift in performance is difficult to achieve in reality. Table 1.1 below summarizes the basic measurements utilized in Earned Value Management.

Table 1

Measurement	Definition
Planned Value (PV)	Budgeted Cost of Work Scheduled
Earned Value (EV)	Budgeted Cost of Work Performed
Actual Cost (AC)	Actual Cost of Work Performed
Cost Variance (CV)	EV – AC
Schedule Variance (SV)	EV – PV
Cost Performance Index (CPI)	EV / AC
Schedule Performance Index (SPI)	EV / PV
To Complete Performance Index (TCPI)	(BAC – EV) / (BAC – AC)

3. Basic EVM Process

The Earned Value Management process begins with determining the project baseline. A list of deliverables that need to be met are captured in a planning process that results in a product-oriented Work Breakdown Structure or WBS (Ruskin, 2005). The WBS is a very important project planning document as it serves as the foundation document for the project. The WBS is what is used to develop an accurate schedule, cost, and staffing plan. For each deliverable listed on the WBS an estimate is needed for how long it will take to complete the task. A time phased budget is created for each activity which rolls up into each work package. The cumulative sum of the individual budgets of the baseline is referred to as the BCWS and includes all relevant project costs to include personnel, material, equipment, and overhead. This data will form into the project baseline plan, or the Performance Measurement Baseline (PMB). Project performance will be measured against the PMB at regularly defined monitoring points established by the organization. At each monitoring point, critical information will be gathered for each deliverable being tracked. For each task being performed, the actual costs should be totaled and the budgeted values for the tasks for which work has been performed should be totaled as well. From these basic pieces of information, the metrics from Table 1.1 are calculated and the health of the project can be understood. As time progresses, the PV, EV, and AC will build up cumulatively. The data from these metrics is typically depicted graphically in an S-curve to quickly and visually communicate project progress.

4. Progress Measurement:

For EVM to properly serve its purpose, a reliable evaluation of the actual status of a project along with timely collection and reporting of data is critical. Progress estimation of a task needs to be as objective as possible and there are several methods suggested in the literature. These cases can be broken down into quantitative and qualitative progressing techniques. The simplest quantitative approach involves counting the number of "units completed" if the project involves production of deliverables that are easily measured. This should primarily be used when something is repeated, and each piece or deliverable requires approximately the same level of effort (Lukas 2008). The "on/off" technique is best utilized when a deliverable cannot be physically measured. In this case, no credit is taken for the task until it is formally accepted as finished and 100% complete. One potential drawback of this is that tasks that are currently underway but not complete are not given any partial credit, so an underestimation of performance could occur. To avoid this issue, "incremental milestones" can be used to credit certain steps of the process with a respective percentage complete associated to it (De Marco, 2009). The "Start – Finish" technique applies some percentage to the start of a task and the remaining percentage upon completion of the task. This is best used when the deliverable is of low value and has a short duration without milestones in between that can be easily defined (Lukas, 2008). The main Qualitative progressing technique is called "Level of Effort" or LOE. This technique is only used when it is very difficult to measure the amount of work that was accomplished. LOE assumes that progress is equal to the actual costs divided by budget (Lukas, 2008). The ability to determine realistic progress for deliverables or work packages is one of the most difficult parts of EVM. Progress should be reported quantitatively as much as possible.

5. Limitations of Earned Value Management:

Based on an extensive literature review, it is clear that Earned Value Management has not been universally accepted or adopted in all projects. Reports of its effectiveness vary throughout, with some reports of positive outcomes (DeMarco, 2009; Chou, 2010) and some reports highlighting the negative outcomes as well (Vargas, 2003; Lukas, 2008). 6 key limitations of Earned Value Management have been identified in the literature which significantly weakens the ability of this Performance Management System to be accepted as a universal best practice across all industries and projects. Each of these 6 key limitations will be discussed further below. Table 1.2 summarizes the strengths and limitations of Earned Value Management that have been identified.

5.1. The Agency Problem

In the defense industry, EVMS is a contractual requirement bestowed upon an organization by the DoD. Inevitability, this leads to a principle-agent relationship as the DoD or "principle" engages with the organization or "agent" to provide some type of service or product on his or her behalf (Jensen and Meekling, 1976). The most important aspect of this relationship is that the authority to make most decisions on the project is delegated by the principle to the agent. Because the principle and agent are from two different organizations, there is a high potential that competing objectives or goals can exist which results in unnecessary costs, delays, and a general mistrust that quickly builds up between the organizations. Further fueling the mistrust between organizations is the fact that information asymmetry (Eisenhardt, 1989) is inevitable as the contractor typically has more information about project issues and progress than the client. Since this asymmetry exists, the contractor may be acting in their own self-interests at the expense of the project, known as the "moral hazard problem" (Turner and Muller, 2004). To address these concerns, principles or contractors have focused on the use of the contract as the unit of analysis (Melnyk, 2004). Namely, a shift to outcome-based contracts has occurred where the principle has information that allows them to verify the claims made by the contracting agency. There has been a recognition that the contract could be used as a mechanism to motivate and control by developing metrics which allow the principle to manage and direct activities to the agent. Even with these metrics, there are still a number of opportunities for information asymmetry to emerge, jeopardizing the relationship between the two organizations and ultimately the ability of the project to be successful. The agency problem discussed above grows when the desires or goals of the principle and agent are in conflict and when it is difficult or expensive for the principal to verify the actions of the agent (Eisenhardt, 1989). The inability to validate the actions of the agent with certainty leads to the problem of risk sharing when principle and contractor have different attitudes toward risk. Each party will behave differently and take different actions based on their own risk preferences. For the contract to be successfully used as a governing mechanism, it should take into account assumptions about the self-interest and risk aversion of all interested parties, any organizational conflict that could be present amongst members, and specify the level of information that is required (Eisenhardt, 1989).

5.2. Assumption that each activity is independent

Earned Value Management considers each cost account to be wholly independent, essentially neglecting any effects of workflow, dependencies, or variability in project control (Kim and Ballard, 2000). It is not uncommon for tasks to follow in a linear sequence and still have an interdependent relationship, especially in an environment where lean practices have been adopted. The following example can illustrate the effects of ignoring workflow dependencies, as seen in Figure 1.1. In scenario 1, let us assume that Task A is a predecessor to Task B. Task A is required to manufacture 100 units of output in one month with a budget of $1 per unit of output or $100 per month. If Task A produces 25 units each week for four weeks, then the workflow is very stable, and Task B is not constrained in any way by the performance of Task A. In this scenario, Task B is completely responsible for its own performance. Now let us assume scenario 2, where Task A is still the predecessor to Task B, however, the output of Task A is highly variable over the four weeks. In this case, assume Task A is still able to produce an output of 100 units in the month, however, it produces 15, 20, 25, and 40 in the four respective weeks of the month. Due to this unreliable workflow, the performance of Task B might be worse if Task B has some type of resource constraint which prevents it from completing more than 30 units in a given week. Current EVM methodology would flag Task B as the issue because it was not able to meet the monthly output requirement of 100 units. In this case though, the inability to meet output requirements of Task B is a direct result of bad performance from Task A. Earned Value Management is not able to reveal that A is the underlying problem due to an unreliable workflow because it assumes that each activity is independent and not affected by any other activity or task.

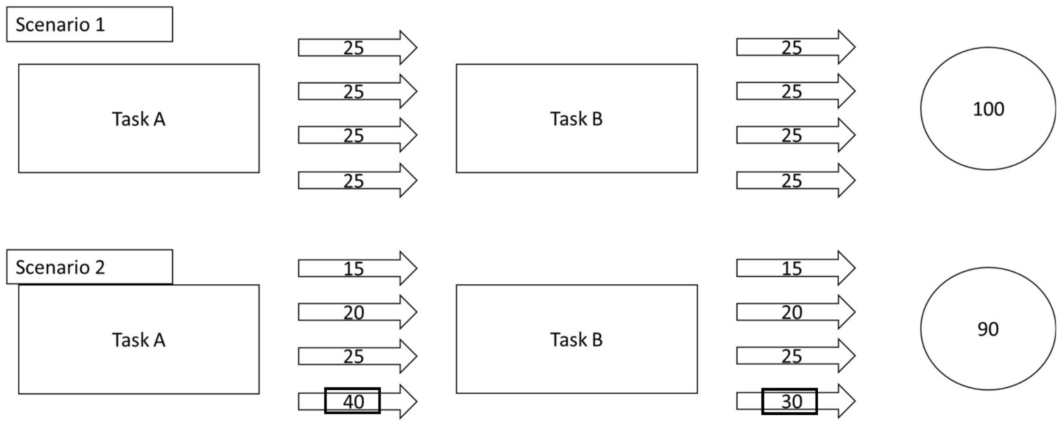


Figure 1

5.3. Assumption that one earned hour is as good as another

Research on cost control in the construction industry has seen tremendous growth over the last decade as more and more construction and development projects are undertaken throughout the world. EVM is an integrated cost and schedule tool where progress is measured based on budgeted dollars or hours to complete specific tasks or work. Specifically, in the construction industry for, example, it is common to assign a budget of x dollars per cubic meter of soil moved (Kim and Ballard, 2002). However, it is clear that a certain sequence must be followed when removing earth. Removing earth from an area that does not need to be at the time is indeed earning value in terms of BCWP, however, it is not actually the one hour of earned value that is needed at that time. This indicates that EVM does a poor job of differentiating between value generating work and non-value generated work.

5.4. Risk of artificially under or over “cooking” performance indexes

As discussed in previous sections, the Schedule Performance Index (SPI) is an overall indicator of whether the project is behind, on, or ahead of schedule. An SPI less than one indicates that work has not been accomplished at the rate in which it was planned. An SPI greater than one indicates that more work has been accomplished than was planned. However, a review of literature indicates that an SPI greater than one does not necessarily mean that a project is ahead of schedule. It is possible that more work can be accomplished than planned by working on non-critical path work packages or by working on “easier” tasks that were scheduled in the future. Basically, if a schedule variance is showing as negative and someone wants to prevent that cost account from showing behind schedule, all that individual would have to do is increase BCWP as much as possible by the reporting date. The development of detailed work sequences is typically left to the discretion of the managers who can sequence and release work to make performance appear better, without regard to workflow or the needs of the following work center. It has been argued that to truly understand the schedule health of a project, project float must also be observed in combination to give a true determination of status (Lukas, 2008). Recently, some scholars have suggested refinements to EVM that would address this issue by integrating activity-based sensitivity information into the calculations of earned value to decrease and ultimately remove false effects from non-critical activities (Elshaer, 2013).

5.5. SV & SPI accuracy

EVM has been a widely adopted method for forecasting project costs. However, the ability to apply EVM to schedule performance has not been as widely adopted due to the standard approach of EVM to evaluate schedule performance in monetary units rather than units of time. The ability to monitor schedule progress and forecast future progress is a powerful tool that EVM intends to capitalize on by utilizing a Schedule Performance Index or SPI. Many researchers have noted the struggle with SPI (Short, 1993; Vandevoorde and Vanhoucke, 2006; Leach, 2005; and Lipke, 2003) which has led to its usefulness being commonly questioned in both industry and academia. The reason for this lies within its fundamental properties. First, like CPI, these performance indices are only accurate for future targets if the project is assumed to run as it has in the past – I.E. no corrective action has been taken which would alter the performance to date. The performance indices strictly assume that projects will proceed as it has in the past. As soon as baselines conditions change, those indices begin to lose their power, and this must be considered when calculating estimates at completion. Secondly, an inherent defect unique to the SPI calculation is that as the project gets closer and closer to completion, the SPI will begin getting closer and closer to a value of 1 even if the project is behind schedule. This is because, upon project completion, the SPI will have a value of 1 as BCWP will be captured for every task on the baseline schedule. As a result, researchers have recommended that SPI only be considered useful during the initial stages of project development (Fleming, 1991). As mentioned earlier, it is important to look at project float along with SPI to confidently identify whether the project is truly behind, on, or ahead of schedule. Thirdly, in EVM, schedule performance is measured, analyzed, and predicted in units of value like money or percent complete as opposed to its natural state which is time. Finally, traditional EVM schedule performance measures have been shown to not be very reliable when the cost of delayed critical activities is a fraction of other cost activities. In this case, the delay in completing the low-cost critical activities will not indicate much of a variance (Khamooshi and Golafshani, 2014).

5.6. EVM does not consider uncertainty or give a range of possible values

A typical control process in project management consists of monitoring actual performance, comparing that with planned performance, identifying variances, and then estimating or forecasting the amount at completion. The main purpose of project control is to provide pertinent stakeholders with the ability to foresee potential future problems and take corrective actions as necessary in a timely manner (Kim and Reinschmidt, 2010). One of the fundamental shortcomings identified in the literature is that EVM based forecasting for both cost and schedule is deterministic and does not provide a range of possible outcomes nor the probability of meeting the objectives of the project (Kim and Reinschmidt, 2010). Without a probabilistic range of outcomes, all possible scenarios cannot be properly considered, and steps cannot be made to plan for varying outcomes. Essentially, the entire team is planning just based off one exact scenario where the probability of that exact outcome is extremely low.

Table 2

Strengths	Limitations
Integrates cost and schedule into a single methodology	Agency problem
Aligns the organization at the strategic and operational level	Neglects workflow, dependencies, and variability in project control
Provides status on current performance as well as projections of future performance	Poorly differentiates between value generating and non-value generating work
Based on a Work Breakdown Structure	Does not consider project float
Enables project managers to gain a better understanding of the true nature of project status midstream	Schedule performance is in terms of monetary units
Promotes disciplined project management planning	Does not give a range of possible values
Early warning indicator	
Historical data captured can be used for comparative analysis in the future	

6. EVM Extensions:

A considerable amount of research has been made related to possible extensions of EVM in response to some of the constraints identified in the previous section. Namely, three main extensions/alternatives to traditional EVM have been identified in the literature which will be discussed in detail below.

6.1. Earned Schedule Management

Earned Schedule is an extension of EVM that was developed in 2003 by Walt Lipke and has now been accepted as a reasonable tool for schedule control. In fact, it is now even included in the appendix in the Project Management Institute (PMI) Standard for EVM. Lipke argued that “from the time of the development of the EVM indicators, it has been known that the schedule indicators are flawed and exhibit strange behavior over the final third of the project when performance is poor (Lipke, 2003)”. The use of Earned Schedule (ES) is rooted in the same fundamentals as Earned Value with one key distinction. ES makes use of schedule performance in terms of time as opposed to cost. ES considers when work was supposed to be completed as opposed to the dollar value of work accomplished. ES was largely developed in response to certain behaviors in traditional EVM that do not allow project managers to fully understand the behavior of their projects in the future. As discussed earlier, SPI converges and concludes at the value of 1.0 so it is difficult to use SPI reliably (Lipke, 2009). There are three main factors that are the basis of Earned Schedule Management (ESM). The first is called the “Planned Duration” or PD, which is the length of time the project task is baselined to take. The second factor is called the “Actual Time” or AT, which is the actual number of time periods that have passed to the measuring point. The third factor is called the “Earned Schedule” or ES and is dependent on the AT. ES is calculated as the time period at which the amount of value earned at AT was supposed to happen per the PMB. For example, let us suppose that the current AT is 6 months and the value earned at month 6 is $100. However, according to the baseline plan, a value of $100 was supposed to be earned at month 4. This would mean that the ES is only 4. With these main factors calculated, it is now possible to derive the Schedule Performance Index (SPI (t)) for ES. This is done by taking the ES divided by the AT. In the example just mentioned, the SPI(t) would be 4/6 or approximately 0.67. The PD divided by the SPI for ES is the equivalent of the EAC that is derived similarly in traditional EVM. In the example discussed, the EAC for ES would be 12/.67

which is approximately equal to 18.18. Lipke conducted a study with nearly 500 months of data that found that ESM analysis yielded higher variation than expected but it consistently had better performance for schedule (Lipke, 2009). Even though SPI (t) has shown evidence of being a superior measurement compared to SPI, it does have some conceptual limitations. The main criticism of SPI (t) is that it measures schedule performance using monetary terms of EV and PV (Khamooshi, 2014). In essence, ES utilizes EV as a means to get to the duration, however, EV has not been proven to necessarily be an accurate measure of schedule progress. Therefore, performance measures that use cost to measure schedule could be inherently weak as a proxy for measuring schedule progress.

6.2. Earned Duration Management

EVM has evolved into its current state through its focus on cost management, control, and financial analysis for many decades (Brandon, 1998; Fleming and Koppelman, 2004; Kim, 2003). However, even as far back as 50 years, there has been a struggle to separate schedule concerns from budgetary concerns (Paige, 1963). In stark contrast to Earned Value and Earned Schedule Management, Earned Duration Management decouples schedule and cost performance measures completely and has developed alternative indices to measure the performance of a project in terms of schedule and cost (Khamooshi and Golafshani, 2013). EVM was primarily developed to focus on the big picture cost and duration within certain stages of a project, however, this has led to a neglect of the buildup of smaller figures of cost and duration. Basically, cost and schedule overruns in one area can be compensated by underruns in other areas. Especially in organizations where different teams or departments are scrutinized for performance, this can be problematic. Earned Duration Management (EDM) makes use of two groups of measures for schedule performance, one set at the micro-level and another set at the macro level.

The following basic notations are used at the Micro Level:

i) Baseline Planned Duration of Scheduled Activity (BPD) – the authorized duration for a task to complete
ii) Planned Duration of Schedule Activity (PD) – the authorized duration assigned to the scheduled work to be completed, the equivalent of PV in EVM
iii) Activity Progress Index (API) – measures the progress of activity.
iv) Earned Duration (ED) – the value of work performed as a proportion of approved duration assigned to it
v) Actual Duration (AD) – the time in calendar units between start of an activity and the reporting period

The API is calculated as the Actual Duration of a task divided by the actual duration plus the estimated duration to complete. The use of this measurement is completely duration based and does not have any dollars associated with it. PD, ED, and AD for Earned Duration Management are the respective counterparts of PV, EV, and AC for EVM. All the basic measures of EDM are dependent on the “time now” or reporting date as well as the authorized BPD in the baseline schedule.

The following basic notations are used at the Macro Level:

i) Baseline Planned Duration (BPD) – the authorized duration assigned to complete the entire project
ii) Total Planned Duration (TPD) – the sum of Planned Duration for all activities to the left of the current reporting period
iii) Total Earned Duration (TED) – the sum of Earned Duration for all activities to the left of the current reporting period
iv) Total Actual Duration (TAD) – the sum of Actual Duration for all activities

6.3. Customer Earned Value

In traditional practices of Lean Methodology, it is of utmost importance to view every situation from the customer's perspective. This philosophy is deeply rooted in many traditional Quality Engineering practices, often referred to as the "Voice of the Customer" or VOC. Rather than focusing on the customer, EVM traditionally focuses inward on outputs and the producer's perspective regardless of what customers need downstream. Customer Earned Value (CEV) was developed by Kim and Ballard to control workflow in an internal supply chain. It defines value from the customer's perspective as to how much of the received work can be used immediately. CEV methodology has defined a list of three major requirements for the application of this methodology (Kim and Ballard, 2002). The first is regarding quality, which states that work should be conformant to all specifications. If the work does not produce a deliverable that meets all specifications, then it is of zero value to the customer. Second, the size or amount of work provided to the customer should be what was agreed upon and baselined to. The customer does not value anything until all agreed upon components and deliverables have been met at the specified quantities. Finally, the customer will only find value in the work that is pulled. The customer should be receiving the work that is ready to be worked on. The main difference between EV and CEV lies within the perspective that each takes, EV takes the performer's perspective whereas CEV takes the customer's perspective.

CEV is defined as the budgeted amount of work completed and used by successors. Any difference between EV and CEV is the level of WIP between the two activities, defined as Value-in-Process, or VIP (Kim and Cho, 2015). The VIP contains the amount of work that was completed but has yet to be used by the successor. CEV is not earned until the work of the previous station is released to use by the next station. The VIP ratio is an indicator of the level of growth or decline in VIP that has occurred from one time period to the next. A value of less than 1.0 indicates less WIP between stations or tasks and a value greater than 1.0 indicates greater WIP between stations or tasks.

7. Concluding Remarks:

Earned Value Management (EVM) is a powerful budgetary and schedule tracking tool that has become widely adopted across all types of projects. The basic principles of this methodology have been utilized since the late 1800s in the industrial world. The first official use of this concept was used in the early 1960s by the United States Air Force for the Minute Man Missile (Vakhrushkina, Mishakova, Borshcheva, 2017)). After the success of the application, this practice became a requirement upon many government contracts. EVM translates work into monetary terms and tracks progress against a baseline plan. By tracking against this point of reference, EVM has the ability to produce a number of metrics that aid in tracking progress as well as predicting future cost and schedule. As discussed earlier, basic metrics such as schedule and cost variance drive performance measures like Cost Performance Index (CPI) and Schedule Performance Index (SPI). From the performance measures, future cost and schedule can be extrapolated to better understand where the project is headed.

Though EVM has been widely adopted on many projects, there are clear limitations indicated in the literature which ultimately inhibit the ability of EVM to be as successful as advertised. There are management shortcomings regarding EVM as it struggles to resolve the agency problem inherent anytime a contractor is acting on behalf of the contracting agency. EVM also assumes that every activity is independent and in many instances is unable to highlight the true root cause of the problem. For those who have a clear understanding of EVM, it is easy to deliberately drive certain actions which can alter the performance measures if they so choose to or are incentivized with certain Performance Based Payments (PBP's). There are rather strong weaknesses inherent in the scheduling side of EVM. SPI will tend to get closer and closer to a value of 1 as the project reaches the end of its life, even if the project is well beyond late to the customer. Not all hours earned are equal in reality, however, EVM treats each hour earned as if it is just as good as another. Finally, EVM will provide an exact estimated cost at completion. The probability that the exact estimated cost will occur is very small. EVM could be more powerful if it could provide a range of possible values as opposed to just one.

In response to some of these shortcomings, researchers have begun looking into alternate methodologies or modifications to traditional EVM. Earned Schedule (ES) is one of these methods. ES considers when work was supposed to be completed as opposed to the dollar value of work accomplished. Earned Duration Management (EDM) has been able to decouple cost and schedule measures for the first time. This is a powerful concept that is a popular area of research today. Finally, Customer Earned Value shifts the focus of performance management from the producer's perspective to that of the customer.

As a result of this extensive literature review, it is clear there are still limitations of EV that exist which prevent it from becoming a universally accepted best practice. Extensions of EV that address the inherent information asymmetry associated with the Agency Problem is a clear area that still needs additional research and insight. In addition, extensions of EV need to be developed which can provide a range of possible values for final project cost and schedule. Earned Schedule Management, Earned Duration Management, and Customer Earned Value have all been developed in response to the remaining limitations brought forth in this paper and summarized in table 1.3 below.

Table 3

Limitation Number	Earned Value Limitation	Earned Value extension addressing the limitation
1	Agency Problem	Research Gap
2	Does not provide a range of possible values	Research Gap
3	Assumption that each activity is independent	Customer Earned Value
4	Risk of "cooking" performance indices	Earned Duration Management
5	SV & SPI accuracy for projects behind schedule	Earned Schedule Management
6	Assumption that one earned hour is as good as another	Customer Earned Value

8. Proposed path forward for Limitation 2:

A proposed methodology to address Limitation Number 2 regarding the inability of EVM to provide a range of possible values is to integrate EVM with traditional Statistical Process Control (SPC). SPC is defined as the use of statistical techniques to control a process or production method (ASQ). This is primarily done through the use of Control charts developed by Walter Shewart in the 1920s. SPC has been applied to project schedule control in four major studies utilizing various control chart methods including CUSUM, Shewhart, and Individual Moving range charts. From these studies, it was found that SPI(t) utilized in the Earned Schedule methodology provides the best option to differentiate between common and special causes of variation. However, the average area under the curve (AUC) for SPI(t) was highest when used directly on the individual moving range control charts. While these studies have laid the groundwork for integrating SPC and EVM, there is still additional research that is needed to further strengthen this methodology. Additional research needs to be done on applying the appropriate transformations of earned value indices, as they are typically non normally distributed. Secondly, the SPC and EVM integration needs to mature to a point where appropriate prediction models of final project performance can be validated and trusted.

Future papers on the integration of SPC and EVM are in process and will look to address the integration of these two methodologies.

9. References

Bryde, D., Unterhitzenberger, C., & Joby, R. (2018). Conditions of success for earned value analysis in projects. International Journal of Project Management, 36(3), 474-484. doi:10.1016/j.ijproman.2017.12.002

Chen, H. L., Chen, W. T., & Lin, Y. L. (2016). Earned value project management: Improving the predictive power of planned value. International Journal of Project Management, 34(1), 22-29. doi:10.1016/j.ijproman.2015.09.008

De Marco, A., Briccarello, D., & Rafele, C. (2009). Cost and schedule monitoring of industrial building projects: Case study. Journal of Construction Engineering & Management, 135(9), 853-862. doi:10.1061/(ASCE)CO.1943-7862.0000055

De Marco, A., Briccarello, D., & Rafele, C. (2009). Cost and schedule monitoring of industrial building projects: Case study. Journal of Construction Engineering & Management, 135(9), 853-862. doi:10.1061/(ASCE)CO.1943-7862.0000055

Earned value project management-fourth edition.(2010). PM Network, 24(12), 69-69. Retrieved from https://login.ezproxy.net.ucf.edu/login?auth=shibb&url=https://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=57268664&site=ehost-live&scope=site

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. Academy of Management Review, 14(1), 57-74. doi:10.5465/AMR.1989.4279003

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics, 3(4), 305-360. Retrieved from https://login.ezproxy.net.ucf.edu/login?auth=shibb&url=https://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=12243301&site=ehost-live&scope=site

Jui Sheng Chou, Hung Ming Chen, Chuan Chien Hou, & Chun Wei Lin. (2010). Visualized EVM system for assessing project performance. Automation in Construction, 19

Khamooshi, H., & Golafshani, H. (2014). EDM: Earned duration management, a new approach to schedule performance management and measurement. International Journal of Project Management, 32(6), 1019-1041. doi:10.1016/j.ijproman.2013.11.002

Kim, Y., & Ballard, G. Is earned value method an enemy of workflow.

Kim, Y., & Ballard, G. (2002). Earned value method and customer earned value. Journal of Construction Research, 3(1), 55. Retrieved from https://login.ezproxy.net.ucf.edu/login?auth=shibb&url=https://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=9278538&site=ehost-live&scope=site

Kim, Y., & Ballard, G. (2005). Profit-point analysis: A tool for general contractors to measure and compare costs of management time expended on different subcontractors doi:10.1139/l05-021

Lipke, W., Zwikael, O., Henderson, K., & Anbari, F. (2009). Prediction of project outcome: The application of statistical methods to earned value management and earned schedule performance indexes. International Journal of Project Management, 27(4), 400-407. doi:10.1016/j.ijproman.2008.02.009

Lukas, J. A. (2008). Earned value analysis - why it doesn't work. AACE International Transactions, , 1-10.

Retrieved from https://login.ezproxy.net.ucf.edu/login?auth=shibb&url=https://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=33720001&site=ehost-live&scope=site

McAdam, R., Hazlett, S., & Galbraith, B. (2014). The role of performance measurement models in multi level alignmentAn exploratory case analysis in the utilities sector. International Journal of Operations & Production Management, 34(9), 1153-1183. doi:10.1108/IJOPM-09-2011-0313

Melnyk, S. A., Stewart, D. M., & Swink, M. (2004). Metrics and performance measurement in operations management: Dealing with the metrics maze. Journal of Operations Management, 22(3), 209-218. doi:10.1016/j.jom.2004.01.004

Ruskin, A. (2005). Variance analysis in earned value measurements when large uncertainties exist in underlying estimates. Ieee, , 4339-4342.

Turner, J. R., & Müller, R. (2004). Communication and co-operation on projects between the project owner as principal and the project manager as agent. European Management Journal, 22(3), 327-336. doi:10.1016/j.emj.2004.04.010

Upadhaya, B., Munir, R., & Blount, Y. (2014). Association between performance measurement systems and organisational effectiveness. International Journal of Operations & Production Management, 34(7), 853-875. doi:10.1108/IJOPM-02-2013-0091

Vandevoorde, S., & Vanhoucke, M. (2006). A comparison of different project duration forecasting methods using earned value metrics. International Journal of Project Management, 24(4), 289-302. doi:10.1016/j.ijproman.2005.10.004

Vargas, R. V. (2003). Earned value analysis in the control of projects: Success or failure? AACE International Transactions, , 1-1. Retrieved from https://login.ezproxy.net.ucf.edu/login?auth=shibb&url=https://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=14698940&site=ehost-live&scope=site

Verma, A., Tahlyan, D., & Bhusari, S. (2018). Agent based simulation model for improving passenger service time at bangalore airport. Case Studies on Transport Policy, doi:10.1016/j.cstp.2018.03.001

Wake, S. (2008). Earned value analysis. Steve Wake Projects Ltd, 8th Edition

Exploratory Analysis of the Malcolm Baldrige National Quality Award Model

Anyama Tettey[1]
Sampson Gholston, PhD[1]
Bryan Mesmer, PhD[1]

[1]*University of Alabama in Huntsville*
aht0005@uah.edu; gholsts@uah.edu; blm0027@uah.edu

Abstract

This paper conducts exploratory statistical analysis to assess trends in scores for the Malcolm Baldrige National Performance Excellence Award Model (MBNQA). The analysis identifies significant differences and similarities across sectors of examiner scores for the award program. The paper makes use of real consensus and site visit scores data collected over 11 years period, 2007-2017, from a States' Performance Excellence Award program to conduct the analysis. There are 2 parts to the study. In the first part, the authors use descriptive statistics and various univariate parametric procedures to observe differences/ similarities in variability and mean scores over the period of the study. The results show that the variability in the scoring approaches across sectors were not statistically different. The mean consensus scores, however, were statistically different from the mean site visit scores. The second part of the study uses multivariate analysis to assess any significant difference in means between award winners and non-award winners for the consensus and site visit scores. The results show a significant difference between award winners and non-award winners for all 7 category scores using the site visit data but a different pattern for the consensus scores. The study confirms examiner consistency and reliability, as well as the need for all applicants to be given a site visit tour during the award program.

1. Introduction

The Malcolm Baldrige Performance Excellence Model has proven to be a valuable tool for use by organizations interested in improving overall Organizational Performance. At a time when the influx of Japanese automobiles and electronics into the American market was seemingly making American companies less competitive, the MBNQA was formed in 1987 to address the ensuing problem. The goal of the MBNQA Improvement Act of 1987 is thus to enhance the competitiveness of U.S. businesses (Link & Scott, 2012).

Since its inception, its usefulness has become apparent to quality experts and other parties interested in the growth of American businesses, and a form of it has been rolled out by most states in the US and all industrialized countries including Japan (Townsend & Gebhardt, 1996). Studies have shown numerous benefits of the award program to the US economy. In 2011, Link et. al, showed that the economic and social benefit to cost ratio had increased considerably from 207-1 for an earlier study they did in 2001 to 820-1 in 2012 (Link & Scott, 2012). Their 2012 study used the ratio of social benefits to social costs for the population of all Baldrige applicants from 2007-10 to come up with the 820:1 ratio in 2012. The authors, however, acknowledged 351:1 as a more conservative ratio since it is more representative of the 45 (16.5%) applicants that responded to their survey.

The importance and usefulness of the model further resulted in the emergence of local and state quality award programs. As of 2016, there were 30 independent Baldrige-based state and regional award programs covering nearly all 50 states (NIST, 2016b).

Despite the success stories for the MBNQA, data used for most research work relating to the Baldrige award has had some obvious limitations. Up until 2009, no data relating to the Baldrige program and scoring process had been made available to the general public due to the sensitive nature of the program. In 2009, however, NIST released the blinded Baldrige scoring data covering years from 1990 to 2006.

Within the period from 1990 to 2006 however, there were some major revisions to the Baldrige framework that makes it unreasonable to identify any observable trends within that stretch of time effectively (Link & Scott, 2012). Within this period the criteria evolved from a quality assurance focus to an overall Performance Excellence focus (Lee, Zuckweiler, & Trimi, 2006) and it was not until 1999 that the scope of the award was increased to include healthcare and education.

This research uses more recent scoring data having both consensus and site visit scores of 59 applicants from a states' MBNQA program from 2007 to 2017. The state has an extensive corps of more than 150 examiners, 40-member board of directors and a panel of 5 judges that make recommendations for award recipients to the state's governor.

The idea of the state programs is to ensure that applicants attain a high level of maturity before applying at the national level. Thus, applicants to the national award program are required to have first applied to the state program unless they apply for a waiver of the standard requirement of first achieving a top-level alliance for Performance Excellence Award.

The national Baldrige scoring procedure involves a joint review of submitted applications by examiners and the subsequent award of scores referred to as consensus scores based on the applicants' achievement in 7 areas known as the Baldrige Criteria for Performance Excellence. At the national level, this process helps to decide whether the applicant will go to the next step of receiving a site visit and subsequently a site visit score. The state award programs closely mimic the national programs with minor modifications, and this is one of such modifications with this states' program where all applicants receive a site visit.

Using the states' data this study tests the following propositions:

a. There is significant difference in scores across sectors, and between site visit and consensus scores
b. There is significant difference in variability across sectors, and between site visit and consensus scores
c. There is significant difference in mean scores to reflect the minor changes in criteria or improvement in applicant performance over the years
d. The significance of each of the seven criteria is the same for determining award winners for the 2 scoring approaches

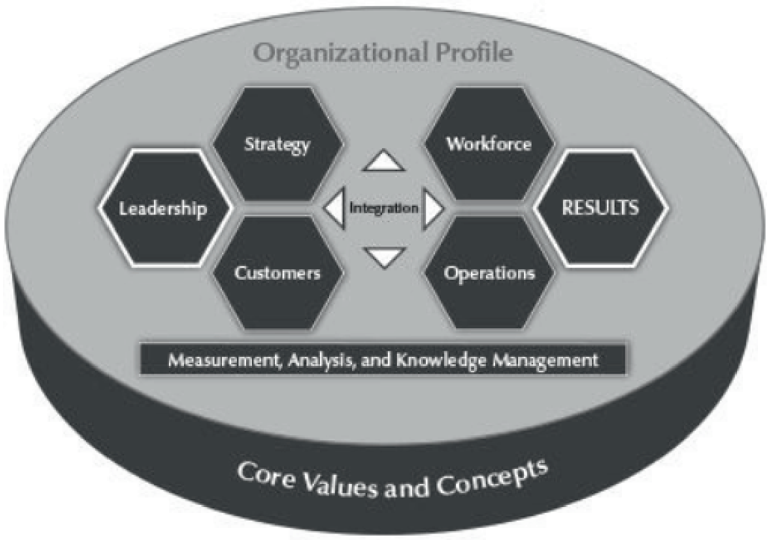
The study concludes with findings that could lead to a useful assessment for all stakeholders involved in or interested in the Performance Excellence agenda in the US.

2. Overview of the Baldrige Excellence Framework and award program

The Malcolm Baldrige National Quality Award (MBNQA) was instituted in 1987 (Steeples, 1994) by the US congress to award United States organizations that have attained a high level of performance excellence based on an assessment by independent examiners.

The importance and usefulness of the award process further resulted in the emergence of local and state quality award programs that ensured that applicants attained a high level of maturity before applying at the national level. Consensus scores are given to applicants based on their answers to questions that represent 7 aspects of the organization. These 7 aspects are used as benchmarks to award scores for applicants in the areas of Leadership, Strategy, Customers, ‘Measurement Analysis and Knowledge Management’ (MAKM), Workforce, Operations and Results (ASQ, 2018). At this state level program, all applicants receive site visits by examiners after their applications have been thoroughly reviewed. At the national level, however not all applicants graduate to the stage of receiving site visits, only a selected few progress to that stage. Again, at the national level, there are 6 different eligible categories for manufacturing, service, small business, nonprofit, healthcare, and education. At the state level whose data is being used for this study however, the same framework and criteria is used for all sectors. One of the motivations for this research is thus to look at any underlying differences in scores that exist between sectors and the two scoring approaches used.

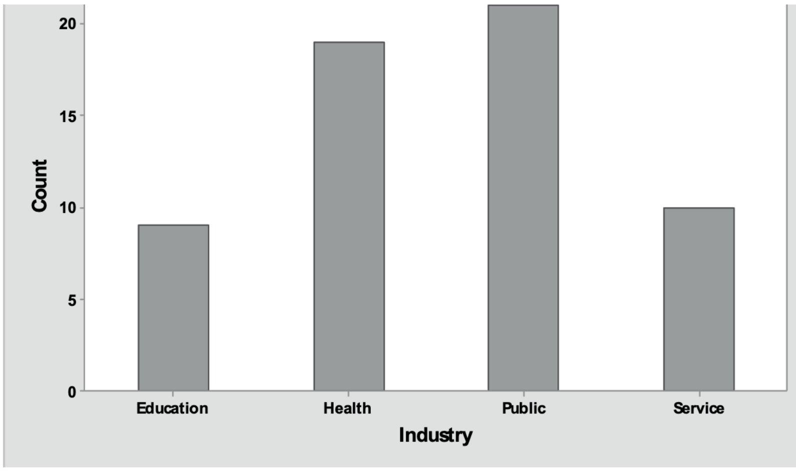
Figure 1 both the national and state a



This state level award program has applicants categorized as healthcare small, healthcare large, service small, service large, education small, education large, public small and public large within this period. For the purposes of this study small and large sectors have been grouped together. The categories are weighted with maximum attainable as: 140, 100, 100, 100, 100, 100 and 360 for each of the categories Leadership, Strategy, Customers, ‘Measurement Analysis and Knowledge Management’ (MAKM), Workforce, Operations and Results respectively, a total maximum attainable score of 1000. After a thorough review of the applicant’s submissions, the total for the 7 categories becomes the consensus score for the applicant. Each applicant then receives a site visit by assigned examiners who visit the organization’s premises. Applicants are rescored after the visit and an applicant is nominated for an award based on the performance. At the national level, up to 18 awards may be given annually across the six eligibility categories. At this State level program however depending on the examiner assessments, there can be as many winners as the number of applicants. Further details on the scoring guidelines and the award national procedure are available at (NIST, 2016a).

3. Descriptive statistics

This section carries out basic descriptive statistics of the data. A brief description of the data is shown in figure 2 below. There are a total of 9 applicants from the educational sector and 19, 21, and 10 applicants respectively from the healthcare, public, and services sectors.



Figures 3 and 4 provide a visual assessment of the trends in mean scores and the variability observed between consensus scores and site visit scores. It can be seen from Figure 3 that the mean scores for the site visits are always higher than those for the consensus scores.

Figure 3 also shows that the pattern across sectors is the same for each of the 2 scoring approaches with the educational sector scores having the lowest means whilst the healthcare sector consistently has the highest means.

Figure 4 shows a box plot of the scores for the four sectors compared over the 2 scoring approaches. The main takeaway from this figure is that there seems to be relatively more variability in the site visit scores compared to the consensus scores for the same sector.

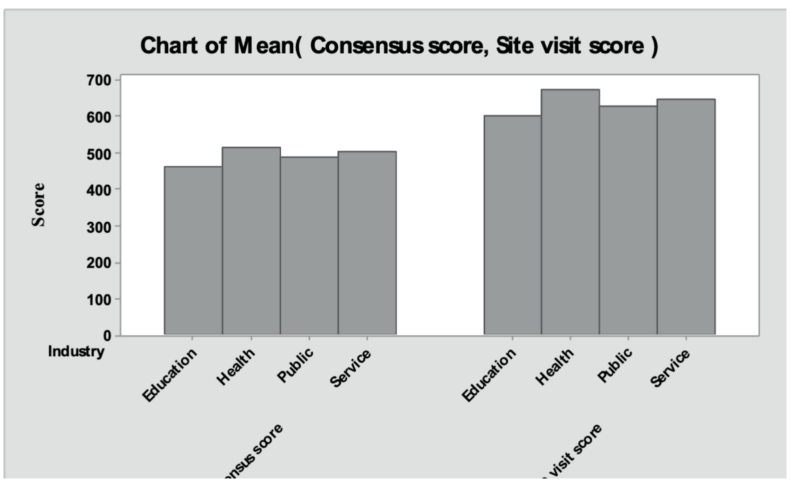


Figure 3. Comparison of mean scores across sectors for the 2 scoring approaches

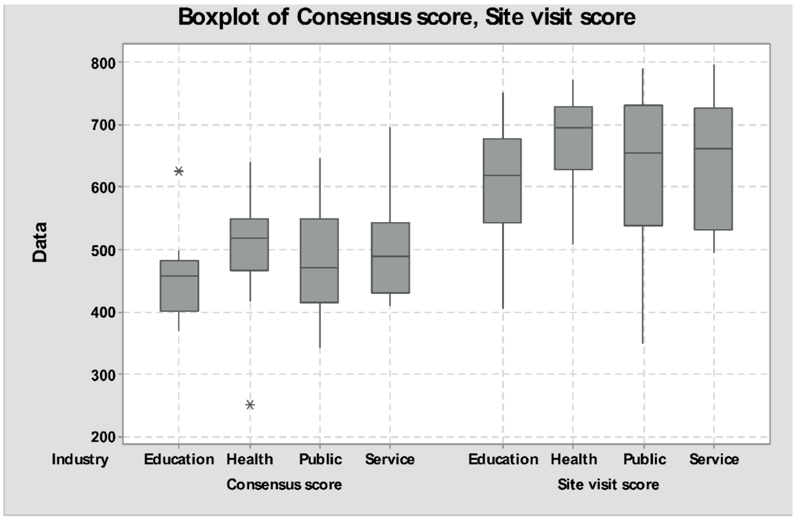


Figure 4. Variability and mean scores for the 2 scoring approaches compared

The authors went ahead to test if the observed differences in figures 3 and 4 were statistically significant using inferential statistical tests. The results are discussed in section 4 below.

4. Inferential analysis

This section uses a 2-sample t-test, one-way ANOVA and posts ANOVA techniques for the analysis. The null and alternate hypothesis of the 2-sample test has $H_0: \mu_{(\text{consensus})}=\mu_{(\text{site visit})}$ and $H_1: \mu_{(\text{consensus})} \neq \mu_{(\text{site visit})}$ respectively. With a p-value < 0.0001 (table 1), we conclude that there is a difference in means for the 2 scoring paradigms. A one-way ANOVA was performed to determine if there is a difference in mean scores across sectors. This was done separately for both scoring approaches. The p-values were 0.465 and 0.351 respectively for the consensus and site visit scores. Thus, in both cases, the conclusion is that the observed differences across sectors (seen in Figure 3) are not statistically significant.

Due to its robustness to the normality assumption, the Levene's test was used for the test of significance of the differences in variability across the 4 sector scores. The test results in Table 1 did not reject the assumption of equal variances. We thus conclude equal variances across all sectors for the consensus and site visit scores within the period 2007-2017. The difference in variability between the site visit and consensus scores was also not significant for both the Levene and Bonett tests as seen in Table 1. Thus all the observed differences in figure 4 were not statistically significant upon investigation.

Table 1. Tests of equality of mean scores and variability within scores

Test method	Result	Equality of the 2 groups (Site visit and consensus scores)	Test method	P-value
95% CI for difference (Mean of consensus - Site visit score)	Interval : -183.4 to -11			
		Levene test of equality across sectors	Consensus	0.714

The third proposition alludes that as minor criteria changes occurred in the years 2007, 2009 and 2013 it reflected in changes in the mean scores over those periods. One-way ANOVA is used to test the equality of means over these periods after a residual analysis showed model adequacy. The null hypothesis is: H_0: $\mu_{(07)}$=$\mu_{(08-09)}$=$\mu_{(10-13)}$=$\mu_{(14-17)}$. The ANOVA results showed a significant p-value and therefore the rejection of the null hypotheses at the 0.05 significant level. Table 2 summarizes the results.

Table 2. ANOVA and Fisher pairwise comparisons across years

Year	Number of applicants	Consensus			Site visit		
		Mean scores	P value	Fisher comparison	Mean scores	P value	Fisher comparison
2007	13	437	0.042	B	573	0.048	B
08--09	14	495		AB	641		AB
10--13	13	517		A	671		A
14--17	19	514		A	670		A

The post ANOVA technique, 'Fisher pairwise comparisons' showed mean differences from 07-10 for both scores. There were no significant differences in mean scores between 2008-2017. Under the column titled 'Fisher comparison' in table 2, years with the same letter show equal means. There is a slight increasing trend from 2007 to 2008 for both the site visit and consensus scores.

5. Inferential analysis using MANOVA

This part of the analysis uses MANOVA to test the hypothesis of equal mean vector scores between award winners and non-award winners for both the consensus and site visit scores. Here we refer to an applicant's category score for all 7 categories, Leadership (x1), Strategy (x2), Customers (x3), MAKM (x4), Workforce (x5), Operations (x6), and Results (x7), (X_j' =x1; x2; x3; x4; x5; x6; x7), as a vector score.

The general case of the MANOVA model is for comparing 'g' (2 or more) population means, with each population consisting of 'p' (more than one) responses (Johnson & Wichern, 2002). We consider the simplest case of comparing 2 population mean vectors: Winners and non-winners, to statistically determine if the 'seven category scores/ responses vary across different levels of the factor. This is done separately for the consensus score and site visit scores. We test the hypotheses that: $H_0: \boldsymbol{\tau_1} = \boldsymbol{\tau_2} = \cdots = \boldsymbol{\tau_g} = \boldsymbol{0}$, where $\boldsymbol{\tau}_\ell$ represents the ℓ^{th} scoring difference; estimated as the ℓ^{th} population mean vector minus the overall sample mean vector.

The statistic Wilks' Lambda corresponds to the equivalent of the F-test in the univariate case. A significant result is followed up by post hoc tests to determine what accounts for the differences. As with all parametric studies, some assumptions needed to be satisfied to pave the way for proper MANOVA analysis.

The following assumptions about the structure of the data for one-way MANOVA were satisfied: The random samples from the 2 populations are independent. The test for the equality of covariance matrix for the 2 populations failed to reject the null hypotheses of equal covariance matrix at a significance level of 0.05. Finally, to assess multivariate normality, Mahalanobis distances within the 7 category scores of an observation, for both the consensus and site visit data did not raise any concerns for alarm. Normality assumption was thus assumed satisfied to carry out a one-way

MANOVA.

The analysis in SPSS gave a significant value of Wilks' Lambda; <0.0001 and 0.031 for the site visit and consensus scores respectively. The result of Fisher's post-hoc 'test of between-subjects effects' is shown in Table 3 below.

Table 3. Results of one-way MANOVA for consensus and site visit score data

Dependent Variable	Consensus score data		Site visit score data	
	F	Sig.	F	Sig.
Leadership	10.5	0.002	25.2	<0.0001
Strategy	8.2	0.006	26.1	<0.0001
Customers	5.7	0.021	16.5	<0.0001
MAKM	4.3	0.044	26.7	<0.0001
Workforce	2.5	0.119	15.1	<0.0001
Operations	2.0	0.160	36.3	<0.0001
Results	14	0.001	72.4	<0.0001

The results in table 3 show a significant difference between award winners and non-award winners for all 7 category scores using the site visit data but not different for all with the consensus scores. Specifically, there was no significant difference between award winners and non-award winners for the workforce and operations category scores.

6. Conclusions

This study explored trends and differences in consensus and site visit scores from a States' Performance Excellence Award Program that have been recorded over 11 years period from 2007-2017, with a total of 59 applicants.

The descriptive statistics showed the educational sector scores having the lowest means whilst the healthcare sector always had the highest means for both the consensus and site visit scores. Scoring trends over the years and across sectors increased slightly in 2008 but not too significantly afterward.

We observe examiner consistency in scoring trends by comparing the site visit data against the consensus data. Criteria changes have not been major within the period of study, and this paves the way for data consistency and validity when performing statistical data analysis with this Baldrige data.

There was no significant difference between award winners and non-award winners for the workforce and operations categories with the consensus scores. There were however significant differences between the 2 sets of scores using the site visit data. This result agrees with what was observed in the study to assess scoring differences between award winners and non-award winners for the MBNQA (Tettey, Gholston, & Mesmer, 2018). These are interesting outcomes that seem to suggest that all applicants deserve to be given a site visit tour.

Further work will investigate the adequacy of theoretical linkages underlying the Baldrige Performance Excellence Model.

7. References

ASQ. (2018). Malcolm Baldrige National Quality Award (MBNQA). Retrieved from http://asq.org/learn-about-quality/malcolm-baldrige-award/overview/overview.html

Johnson, R. A., & Wichern, D. W. (2002). Applied multivariate statistical analysis (Vol. 5): Prentice hall Upper Saddle River, NJ.

Lee, S. M., Zuckweiler, K. M., & Trimi, S. (2006). Modernization of the Malcolm Baldrige National Quality Award. International Journal of Production Research, 44(23), 5089-5106. doi:10.1080/00207540500161043

Link, A. N., & Scott, J. T. (2012). On the social value of quality: An economic evaluation of the Baldrige Performance Excellence Program. Science and Public Policy, 39(5), 680-689. doi:10.1093/scipol/scs052

NIST. (2016a). 2017-2018 Baldrige Excellence Framework and Criteria. Retrieved from https://www.nist.gov/news-events/news/2016/12/2017-2018-baldrige-excellence-framework

NIST. (2016b). Four U.S. organizations receive nations highest honor for Performance Excellence. Retrieved from https://www.nist.gov/news-events/news/2016/11/four-us-organizations-receive-nations-highest-honor-performance-excellence

NIST/BPEP. (2018). How Baldrige Works Retrieved from https://www.nist.gov/baldrige/how-baldrige-works

Steeples, M. M. (1994). The Baldrige Award and ISO 9000 in the quality management processes. IEEE Communications Magazine, 32(10), 52-56. doi:10.1109/35.329025

Tettey, A., Gholston, S., & Mesmer, B. (2018). ASSESSING SCORING DIFFERENCES BETWEEN AWARD WINNERS AND NON-AWARD WINNERS FOR THE MALCOLM BALRIGE NATIONAL QUALITY AWARD.

Townsend, P., & Gebhardt, J. (1996). The importance of the Baldrige to US economy. The Journal for Quality and Participation, 19(4), 6.