DART: A Data Analytics Readiness Assessment Tool for Use in Occupational Safety

A Thesis
by
MAIRA COMPAGNONE

Submitted to the School of Graduate Studies
at Appalachian State University
in partial fulfillment of the requirements for the degrees of
MASTER OF ARTS and
MASTER OF BUSINESS ADMINISTRATION

May 2020
Department of Psychology

DART: A DATA ANALYTICS READINESS ASSESSMENT TOOL FOR USE IN OCCUPATIONAL SAFETY

A Thesis
by
MAIRA COMPAGNONE
May 2020

APPROVED BY:

Timothy Ludwig, PhD.
Chairperson, Thesis Committee

Shawn Bergman, PhD.
Member, Thesis Committee

Yalcin Acikgoz, PhD.
Member, Thesis Committee

Rose Mary Webb, PhD.
Chairperson, Department of Psychology

Mike McKenzie, Ph.D.
Dean, Cratis D. Williams School of Graduate Studies

Abstract

DART: A DATA ANALYTICS READINESS ASSESSMENT TOOL FOR USE IN OCCUPATIONAL SAFETY

Maira Compagnone
B.A., University of Massachusetts, Boston
M.A. Appalachian State University
M.B.A. Appalachian State University

Chairperson: Timothy Ludwig, PhD.

The safety industry is lagging in Big Data utilization due to various obstacles, which may include lack of analytics readiness (e.g. disparate databases, missing data, low validity) or competencies (e.g. personnel capable of cleaning data and running analyses). A safety-analytics maturity assessment can assist organizations with understanding their current capabilities. Organizations can then mature more advanced analytics capabilities to ultimately predict safety incidents and identify preventative measures directed towards specific risk variables. This study outlines the creation and use of an industry-specific readiness assessment tool. The proposed safety-analytics assessment evaluates the (a) quality of the data currently available, (b) organizational norms around data collection, scaling, and nomenclature, (c) foundational infrastructure for technological capabilities and expertise in data collection, storage, and analysis of safety and health metrics, and (d) measurement culture around employee willingness to participate in reporting, audits, inspections, and observations and how managers use data to improve workplace safety. The Data Analytics Readiness Tool (DART) was piloted at two manufacturing firms to explore the tool's reliability and validity. While there were reliability concerns for inter-rater agreement across readiness factors for individual variables, DART users agreed on and accurately assessed organizational capabilities for each level of analytics.

Acknowledgments

I must foremost thank Dr. Tim Ludwig, my committee chair, who was paramount in the success of this project, and deserves my deepest gratitude. He provided structure and guidance as I began to explore this area of research and has provided me with every opportunity to expand my knowledge within safety through the classroom, mentorship, and exposure to his work. I owe special thanks to Dr. Ludwig for challenging me and shaping my skills as a researcher.

I must also extend thanks to the other members of my thesis committee. I have had the distinct pleasure of working with my committee members on various undertakings outside of this thesis. Dr. Shawn Bergman and Dr. Yalcin Acikgoz have given me every opportunity to grow as an academic and a professional, and my success today is due in large part to how committed they each have been to my development.

This project was much improved because of the knowledge of experts within safety and industry. I would like to thank Dr. Anne Foreman, Dr. Oliver Wirth, and Dr. Jon Friedel from the National Institute of Occupational Safety and Health for lending their expertise to this project. Steven Addington, Jeff Arrington, Dr. Angie Lebbon, and Al Gourley, representatives from Eastman Chemical Company, provided access to organizational subject matter experts, guidance on various analytics projects, and feedback on this project. Finally, Cory Ammons, Jimmy Summers, Michael Scott, and Amber Friday, representatives from Elevate Textiles, gave this project life with their enthusiasm, cooperation, commitment, and fast turn-around of deliverables and feedback throughout the process.

This project could not have come together without the efforts of the HR Science Safety Analytics Team. The entire team participated in site visits, biweekly meetings with participating organizations, cleaned and aggregated data for our participating organizations, and provided input and feedback on the DART as it morphed and took shape. I also need to extend an extra special thanks to Tara O'Neil, who kept me sane and on task as we attempted to create a concrete product out of an idea. She contributed countless hours, late nights, early mornings, effort, and friendship, even when we were both exhausted and overcommitted to other projects. I appreciate her ability to make me laugh no matter how much stress I am feeling, and how she has supported me throughout every stage of this project.

I truly believe that my graduate cohort is something special, and I must thank everyone in the class of 2020. We have all gone through personal and professional hardships over the last year, but have met these challenges with grace and integrity with each other's support. The everyday informal chats, cohort trips, game and movie nights, and camaraderie I have experienced over the last two years are immeasurable. From the bottom of my heart, I must extend my warmest thanks to Darby Ford, Bill Griffin, Jessie Harris, Philip Hinson, Soundarya Kanthimathinathan, Matthew Laske, Erin Marion, Brooke Myall, Tara O'Neil, Rosalyn Rease, Kelly Redd, and Lindsey Wuerfel, for being a part of my life.

Finally, I would like to acknowledge Matthew Laske. I don't know if I would have undertaken or succeeded in this project if not for his wisdom, knowledge, and support. There is no person I would rather argue about scientific theory with, and no one who's opinion I would trust more on the finer points of APA style or correct citations. More importantly, beyond the scope of this research, I must thank him for always listening to my dreams and finding a way to support me as I reach out for them, for pushing me to be a better person, and for being my best friend.

Dedication

This thesis is dedicated to my dad, who inspires my work in safety. I grew up hearing stories about his job; sometimes funny stories about his coworkers, and sometimes sad or scary stories about accidents and injuries. It is my sincerest hope that my work can contribute to keeping someone else's family member safe.

Table of Contents

Abstract....................iv

Acknowledgments....................vi

Dedication....................viii

List of Tables....................xi

List of Figures....................xii

Foreword....................xiii

Introduction....................1

Creation of the DART Assessment and its Use....................11

Pilot Study....................34

Pilot Study Results....................40

Discussion....................62

References....................69

Appendix A. Initial Meeting with Host Organization....................76

Appendix B. Measurement Framework....................82

Appendix C. Variable List....................83

Appendix D. DART Excel © Workbook....................86

Appendix E. Company B Sample Data Excel © Sheet Cataloguing Available Data....................93

Appendix F. Measurement Culture Survey....................94

Appendix G. IRB Approval....................97

Appendix H. DART PDF Manual....................98

Appendix I. Measurement Culture Survey Summary Statistics....................112

Appendix J. DART Reports for Recommendations (Companies A & B)...................... 113

Vita 118

List of Tables

Table 1. Summary of DART Readiness Factors with their definitions 25

Table 2. Rating Criteria for Each Maturity Stage of the Organization-Level Readiness Factors 27

Table 3. Rating Criteria for Each Maturity Stage of the Variable-Level Readiness Factors 29

Table 4. Scoring System for Readiness Factors: Weights and Minimum Required Ratings to Achieve levels of Analytic Maturity 32

Table 5. Summary of Company A's DART Average Readiness Factor Ratings by Respondent, Averaged Across all 23 Variable Ratings from the Variable List 42

Table 6. Comparison of Company A's DART Average Readiness Factor Ratings to the Research Team Ratings 43

Table 7. Summary of DART Optimization Scores for Each Level of Analytics Per Respondents at Company A 44

Table 8. Inter-rater Agreement (IRA) for Readiness Factor Scores Across all 23 Variables in the Variable List 45

Table 9. Agreement Across Average Readiness Scores for all DART respondents versus the Research Team 46

Table 10. Summary of DART Results Company B 47

Table 11. Comparison of Company B's DART Average Readiness Factor Ratings to the Research Team Ratings 48

Table 12. Summary of DART Optimization Scores for Each Level of Analytics Per Respondent at Company B 49

Table 13. Inter-rater Agreement (IRA) for Readiness Factor Scores Across all 23 Variables in the Variable List for each Company B Respondent 50

Table 14. Agreement Across DART Ratings from Company B Corporate Respondent 1 to Research Team Respondents 52

Table 15. Agreement Across Average Readiness Scores for all DART respondents versus the Company Respondents versus the Research Team 53

Table 16. Comparison Between DART Culture Self-Assessment and Survey Results 54

Table 17. Means, Standard Deviations, and Correlation of Company A Variables 58

Table 18. Variable List 83

Table 19. Measurement Culture Survey 94

Table 20. Culture Survey Descriptive Statistics for Company A 112

Table 21. Culture Survey Descriptive Statistics for Company B 112

List of Figures

Figure 1. DART Data Collection and Discovery Process 24

Figure 2. Research process 36

Figure 3. Graph Displaying the Cumulative Monthly Accident Rate by Fiscal Year and Plant 57

Figure 4. Average number of monthly accidents each fiscal year for each plant 61

Figure 5. Measurement Framework 80

Foreword

This thesis is written in accordance with the style of the Publication Manual of the American Psychological Association (6th Edition) as required by the Department of Psychology at Appalachian State University.

DART: A Data Analytics Readiness Assessment Tool for Use in Occupational Safety

Maira E. Compagnone

Appalachian State University

DART: A Data Analytics Readiness Assessment Tool for Use in Occupational Safety

When employees interact with hazards on the job, there is the risk for potential safety incidents resulting in injury or death. Despite the steady decline of incidents since 1992, when the injury rate per 100-full time employees was 8.9 compared to 2.6 in 2017, safety incidents still have a major impact on the workforce (United States Department of Labor [USDL], Bureau of Labor Statistics [BLS], 2018a). In 2017 there were approximately 2.8 million injuries and illnesses within the private work sector in the United States (USLD, BLS, 2018a). Costs associated with these work injuries goes beyond workers' compensation. The National Safety Council (2019) estimates that work injuries cost companies $161.5 billion, from expenses such as wage and productivity losses ($50.7 billion); medical expenses ($34.3 billion); and administrative expenses ($52.0 billion).

The consequences of workplace incidents also translate to lost time for the organization. In 2017, workplace injuries and illnesses led to 70 million days away from work, and it is estimated that 55 million days will be lost in subsequent years from those 2017 injuries (National Safety Council, 2019).

Since 1970, when the Occupational Safety and Health Administration (OSHA) was created in the Occupational Safety and Health Act (OSH Act), workplace injuries and fatalities have decreased by over 65 percent (OSHA, 2012b). In 1970, an average of 38 workers were killed every day on the job; a number that has since fallen to about 14 workers per day. In recent times, however, the rate at which fatalities have decreased is slowing, with 6,217 fatalities occurring in 1992 compared to 5,147 fatalities occurring in 2017 (USLD, BLS, 2018b). As another example, after 25 years of steadily improving rates in construction safety, fatalities began increasing again at alarming speed; fatalities almost doubled in New

York City between 2014 and 2015 (Ringen, Dong, Goldenhar, & Cain, 2018). This trend signifies that much still needs to be done to determine why fatalities are not being reduced to the same extent injuries are.

Initiatives by OSHA (2012a) and other regulatory bodies have done much to reduce injuries and illnesses, through regulations such as (a) providing information to all employees who may come in contact with hazards or chemicals, (b) providing adequate training, (c) providing no-cost personal protective equipment to all employees, and (d) forbidding retaliation against those reporting unsafe conditions. In addition, organizations are required to notify OSHA when injuries and fatalities occur in the workplace, and store large amounts of data that can be presented as proof of compliance with these regulations.

As such, many industrial organizations have large amounts of data available that could be analyzed statistically. Departments such as supply chain management and production have already gleaned value by using analytics, but safety departments are only just beginning to tap into this resource. “Big Data,” or data that are so large, fast or complex that it’s difficult or impossible to process using traditional methods (Statistical Analysis Software [SAS], 2019), is becoming commonplace in industries globally. Current analytics research within occupational health and safety (OHS) has demonstrated the predictive capabilities of (a) demographic information, such as age, gender, and worker experience (Chi, Lin, & Dewi, 2014; Stewart, 2013), (b) job-related information such as industry, equipment, job risk, and training (Lingard, Hallowell, Salas, & Pirzadeh, 2017), and (c) behavioral information, such as the use of personal protective equipment, hazard identification, and housekeeping (Mistikoglu et al., 2015) in predicting adverse safety outcomes.

By incorporating 112 million safety observations and safety incident data from over 15,000 work sites, safety analytics predictive models developed by Carnegie Mellon University (*Predictive Analytics in Workplace Safety: Four "Safety Truths" that Reduce Workplace Injuries*, 2012) predicted incidents at actual worksites with accuracy rates as high as 80-97 percent, with a high degree of correlation between predicted and actual incidents (i.e. r^2 as high as .75). Using these predictive models, organizations can direct scarce resources to the locations and work teams that are at the highest risk of having safety incidents (Schultz, 2012). Deloitte, a private consulting firm, analyzed five years of data from Goldcorp, a gold-mining company, containing information on injuries, demographics, production, operations, and weather. Relationships between injury rates and compensation, age, job roles and other operational factors were demonstrated (Stewart, 2013). Goldcorp used these insights to increase managerial training, write new policies, and focus supervisor support on employees with higher risk profiles. With this improved and targeted decision-making, Goldcorp hopes to reduce injuries and fatalities.

Research has shown that firms with perceived higher levels of analytics maturity perform better (Cosic, Shanks, & Maynard, 2012), however, not all firms that have made large investments in analytics have achieved improvements in performance and value. The use of data analytics within occupational health and safety has been lagging behind other industries due to many unique challenges. One issue is that large amounts of safety data are reliant on worker observations and voluntary reporting. In many cases employees are unwilling to report on their own or coworkers' risks or take the time to identify hazards. As a potential solution, within recent years, the safety industry has begun to explore machine learning (Vallmuur et al., 2016) or Hadoop data management of automatically captured photos and videos (Guo, Ding, Luo, & Jiang, 2016). Despite these advanced tools, organizations have been slow to run advanced levels of analytics (Hadaya & Pellerin, 2010) because of obstacles beyond reporting, including disparate data platforms, data input errors, or the lack of support staff to organize, clean, and run statistics on the data (Gao, Koronios, & Selle, 2015).

In order to maximize the potential for analytics in OHS, organizations need to be ready to assess their current capabilities for analytics and facilitate meaningful change that will develop those capabilities until they are optimal. Determining current capabilities will allow organizations to utilize the most advantageous level of analytics to inform decision-making.

Levels of Analytics

There are four levels of analytics that differ from each other in terms of their ability to drive improved decision-making and the degree of sophistication of the recommended solutions (Fred & Kinange, 2016): (a) descriptive analytics, (b) diagnostic analytics, (c) predictive analytics, and (d) prescriptive analytics (Lepenioti, Bousdekis, Apostolou, & Mentzas, 2020).

Descriptive analytics answer questions about what has happened in the past. Within OHS, safety data and information is analyzed for characteristics and relationships through descriptive statistics and data visualizations (Huang, Wu, Wang, & Ouyang, 2018). Examples of this type of analytics include sums, means, and averages used to clarify and define the current safety state of an organization through reports and dashboards. In OHS, descriptive analytics may look at the number of behavioral observations or equipment inspections in a month and the distribution curve across the different departments in the company.

Diagnostic analytics provide clues about the reason for such past occurrences. These types of analytics (i.e., correlational analysis) describe relationships between variables to provide context. Diagnostic analytics use historical and past safety performance to identify reasons for the success or failure of initiatives, or reasons for specific safety outcomes by investigating relationships, outliers, and sequences (Huang et al., 2018). Safety data analytics at this level has focused on correlates to workplace injury such as external pressures, internal social context, and organization characteristics including job demands (e.g. environmental conditions, scheduling and workload, physical job demands, and the overall complexity of work; Barling, Loughlin, & Kelloway, 2002), leadership (e.g. relationship with the manager, leadership style, trust, and accountability; Fogarty, 2004), and organizational commitment to safety (Fogarty, 2004).

Predictive analytics (i.e., regression analysis) attempts to prognosticate future outcomes to answer what incidents may happen and why. In addition to historical data, predictive analysis incorporates current information in an attempt to predict the likelihood of an event occurring (Huang et al., 2018). Predictive analytics in OHS can include both short-term and longer-term predictions. For example, Lingard, Hallowell, Salas, and Pirzadeh (2017) used injury rate and other data from a large construction company's safety program (e.g., toolbox talks, prestart meetings, safety observations, hazards reported, etc.) to identify a lagging predictive pattern between indicators. Their findings suggested that injuries were often followed by an increase in preventative measures (e.g. toolbox talks) which then decreased until the occurrence of the next injury. In another example, a construction contracting company in Singapore used five different types of machine learning models on project-related and safety variables to predict conditions under which no accidents would be likely to occur, conditions which may cause minor accidents, and conditions which may cause major accidents. The most effective of these, a random forest regression, a type of decision tree, predicted these conditions with 78% accuracy (Poh, Ubeynarayana, & Goh, 2018).

Finally, prescriptive analytics take results from predictive analytics and utilizes real-time data streams in order to provide the most accurate guidance for decision-making (Mousanif, Saba, Douiji, & Sayad, 2014). Safety data, mathematical formulae, safety rules, and machine learning are analyzed with constantly updated models to suggest the most advantageous real-time decision options based on the identification of future opportunities or risks (Huang et al., 2018). For example, Ayhan, Costas, and Samet (2018) built a Viterbi algorithm variant (i.e., a dynamic programming algorithm that finds the most likely of events within a sequence) to predict when an airplane trajectory may infringe on the protected zone of another aircraft, in order to reduce the chances of air traffic and collisions. The prescriptive model detects and resolves these potential conflicts before aircraft even depart, resulting in safer operations, higher efficiency, and higher capacity, reducing air traffic controller workload.

The levels of analytics require different levels of data characteristics. For example, prescriptive analytics requires data that updates in real-time, while descriptive analytics can use archival data to assess past trends. Organizations must have an accurate picture of their data capabilities in order to focus on the level of analytics that will provide the most accurate and useful output to guide decision-making. Maturity models are one method for determining an organization’s current capacity for analytics, while data readiness assessments provide diagnostic information for improvements that can be made to help organizations optimize and move to higher levels of analytics.

Analytics Maturity Models and Data Readiness

Maturity models assist organizations with assessing current capabilities. The maturity of a process is determined across contiguous stages, from initial maturity to optimized. Readiness assessments apply the framework of maturity stages to multiple components of a process, rather than assessing the maturity of the process in its entirety. Readiness assessments provide information about the specific components of the process that must be improved, thus promoting the development of analytics maturity in organizations who wish to improve from a lower to a higher level of analytics maturity, maximizing the value to be gained.

Maturity models. Maturity models describe typical patterns in analytic process development in organizations as they implement new technologies or capabilities attempting to attain higher levels of analytics for their predictive and prescriptive qualities (Comuzzi & Patel, 2016). Analytics maturity models describe (a) an organization’s current data needs, (b) the analytics that the organization is capable of performing currently, and (c) how to improve processes to achieve advanced levels of analytics (Cosic et al.,2012).

The Capability Maturity Model (CMM; Pfleeger, 1995) provides the framework for most maturity models within information technology, software, and data processes, including an analytics adaptation (Analytic Processes Maturity Model (APMM); Grossman, 2018) and a data management adaptation (Master Data Management Maturity Model (MD3M); Spruit & Pietzka, 2015). The CMM assesses analytical readiness through the availability of inputs (e.g. data variables), the quality of outcomes (e.g. relationships like correlations or regression coefficients with validity and reliability), and the overall establishment of a consistent process that would produce the same results even when initiated or completed by different experts. Specifically, the CMM assesses analytics maturity along five possible stages: (a) initial, (b) repeatable, (c) defined, (d) managed, and (e) optimizing (Pfleeger, 1995).

The first stage (Initial) is a baseline categorized by ill-defined inputs. A lack of definition and consistency means that the data variables can only be loosely connected to expected outcomes, which causes wide variety in the quality of information gleaned from analyses (Pfleeger, 1995). At this level, when organizations are just starting to explore what can be done with current capabilities, analytic results are difficult to explain and interpret (Comuzzi & Patel, 2016).

At the second stage of maturity (repeatable), process management begins. Input variables like dates, outcomes, and constraints (e.g. budget and time) are clearly defined (Pfleeger, 1995). Essentially, though the process by which data inputs create statistical outcomes is not yet explicit, different analyses could be run on the same set of inputs and yield similar outcomes.

At the third stage of maturity (defined) the data analytic process has defined activities (Pfleeger, 1995). Activities refer to the different stages of analytics, including collecting metrics, defining variables, designing systems and code, and testing results. Defined activities increase the consistency and repeatability of analysis, and increase efficiency. At this stage, preliminary assessment of the quality of the outcomes should begin. Deviations from the planned or expected outcomes may shed light on problems with metrics, methods of collecting data, or coding.

Stage four of maturity (managed) has additional oversight, and the entire data analytic process is evaluated for effectiveness. Information and feedback from earlier activities allow for redesign and can provide insight for where resources should be allocated to improve the system. Additionally, because there is information on the whole process (e.g. input, analyses, and outcomes), direct effects from changes in a singular activity can be assessed for efficiency and improved accuracy.

The most advanced stage of data maturity (optimized) leads to the highest level of data analytic processes. At this level, the process is able to improve upon itself through system feedback. Activities may be added or removed automatically, using established algorithms or machine learning, in response to measurement and outcomes. (Pfleeger, 1995).

The stages of maturity are nested, meaning that at the optimal maturity, all lower stages are considered optimized. In addition, the stages may not be discrete, but rather should be considered a continuum of assessment across many dimensions, of which different components may be missing, clearly defined, or more mature at each nested level (Pfleeger, 1995). As such, to make self-assessment easier for a rater with average analytics knowledge and expertise, this study will look at three maturity stages: (a) low maturity, which corresponds to the initial to repeatable stages of the CMM, (b) average maturity, which corresponds to the defined stage of the CMM, and (c) optimal maturity, which corresponds to the managed to optimized stages of maturity in the CMM.

Readiness assessments. Readiness assessments use maturity models to assess an organization's current stage of maturity (i.e. sophistication of data management such as data collection, quality, storage, and manipulation), but use the additional framework of individual components (Klievink, Romijn, Cunningham, & de Bruijn, 2017) to provide recommendations for how to improve readiness. Other industries have developed tools to assess data readiness, including healthcare (Snyder & Fields, 2006), education (Arnold, Lonn, & Pistilli, 2014), and supply chain operations (Nemati & Udiavar, 2013). Although general readiness assessments provide specificity by breaking down organizational analytics maturity into different components that can be individually assessed, industry-specific readiness assessments can provide additional value by aligning the focus of the assessment to variables and relationships found to be of importance in the industry.

Snyder & Fields (2006) developed a data readiness assessment in the healthcare industry to assist in the utilization of predictive variables to create safer environments, specifically by reducing medication errors and adverse drug outcomes. Also within healthcare, the Healthcare-Analytics Pre-Adoption Readiness Assessment (HAPRA) Instrument (Venkatraman, Sundarraj, & Mukherjee, 2016) guides organizations to self-rate their maturity across medical technologies, IT, user adoption of technology, data quality, and management. After completing the assessment, the user is given a readiness score along with advice on how to improve. Within supply chain operations, analytic readiness assessments have also been useful to prepare data structures for analytics to improve efficiency, quality, and supply chain strategies (Nemati & Udiavar, 2013).

Although readiness assessments have been developed and implemented in other industries, research has yet to develop tools to assist OHS in identifying current capabilities for big data analytics using safety antecedents and outcome variables.

Creation of the DART Assessment and its Use

The safety industry often does not have all of the necessary components in place to take advantage of higher levels of analytics that can lead to the prediction and mitigation of injuries. While many organizations may be interested in the valuable outcomes safety data analytics might provide, they may not know if their safety measurement systems are adequate for analytics nor how they might build their capability to mature these systems. Therefore, an analytics readiness assessment is needed for organizations to be able to reliably predict their level of success at analytics (i.e., descriptive, diagnostic, predictive, prescriptive; (Lepenioti et al., 2020) and be able to determine the key components of data collection and storage that need to be further developed. Accordingly, the first purpose of the current study is to create a functional analytics readiness assessment tool for the safety industry.

Data Analytics Readiness Tool (DART) Components

The **Data Analytics Readiness Tool (**DART) was created with a foundation on theory and literature in occupational health and safety, incorporating an assessment of four factors (e.g. Arnold, et al., 2014; Cozummi & Patel, 2016; Eybers & Hattingh, 2017) that have been shown by prior research to influence safety analytics outcomes. These factors include data components of readiness (i.e. data quality) as well as organizational components of readiness (i.e. rules and operations, foundational infrastructure, and safety measurement culture).

Data components of readiness. Assessing an organization's data is foundational to a readiness assessment. Data collected must be of quality, meaning it has validity, reliability, and variability, but in order to run advanced analytics, the data must also meet other requirements. Organizations must assess the types of available data (e.g., observation checklists, video feed, photographs, reports) as well as the volume of data and the speed at which data is updated. Mature organizations that are ready for advanced safety analytics likely collect large amounts of data on a daily basis (McSween, 2003).

Data Quality. High-quality data are foundational for analyzing and using data to realize value (Cai & Zhu, 2015). Without quality data, analytic statistics can still, at times, be run but the results are likely to be subject to Type 2 error where key relationships will not be found. Unfortunately, archival databases of organizational safety measures are often created for a purpose other than conducting analytics. The quality of the data is driven by the original use of the measurement process. Many safety processes like inspections collect large amounts of data but are only used to find equipment issues to fix. Human resources employment data, such as overtime, are collected for payroll but not used to assess the amount of time an employee is working and how it may relate it to injuries. In many cases like these, the measurements were not designed to satisfy the quality requirements for the proposed analyses. Accordingly, the DART assesses the quality of the data for its adequacy to conduct different levels of initial and advanced analytics.

Recognizing that different sources of data can be at various levels of quality, all individual variables across the Variable Matrix are assessed for quality. The evaluation of quality in the DART assesses data validity/accuracy, reliability, and variance.

*Validity/accuracy***.** Validity assesses the extent to which inferences from the data accurately represents the "real world" phenomenon targeted by the measurement (Sattler, 2016). Often, validity is of concern in the safety industry because many metrics are reliant on employee reporting, which may be affected by culture (the moderating effect of measurement culture will be discussed in the organizational components section) or biases (Salas & Hallowell, 2016).

Reliability. Reliability refers to consistency of measurement across time and units. Disparate databases and collection methods can lead to differing variable nomenclature, which may lead to decreased consistency. In addition to variables being named and defined in the same manner across systems, consistency refers to the format and scaling of a variable. For example, data collected on "overtime" may be consistent in that each measure contains information about time spent at work beyond a predetermined schedule of time, but one department may collect information on the minutes worked beyond eight hours in a day, where another department may collect information on the hours worked beyond a total of forty, whether that begins on the third day of the workweek or the fifth.

*Variability***.** Finally, data must contain enough variance to conduct statistical analyses. Variance is defined as the distance of a measure from the mean of the total sample of measures. Variance is commonly measured in statistics as a standard deviation. In a normal distribution, 68% of the measures are typically located within one standard deviation from the mean, 95% of measures within two standard deviations, and 99% of all measures are located within three standard deviations from the mean.

If range restriction, or truncated variance with a very low standard deviation, occurs in a single variable, causing an abnormal distribution shape (e.g. leptokurtic distribution shapes contain greater amounts of measures very close to the mean), then the chance of finding a correlation with any other variable becomes limited (Type 2 Error). Many statistical analyses based on the general linear model rely on adequate variance in variable distributions in order to find relationships. This is the case with most analytical statistical techniques.

A number of variables in safety, such as the main outcome variable of injuries, naturally have very little variation due to the low frequency of occurrence. In addition, there may be restrictions on the variance of predictor and precursor data because measures require individuals to voluntarily recognize the event(s) and record/enter the data. For example, the reporting of minor injuries, close calls and at-risk behaviors may be truncated because workers don't perceive their importance, forget to stop their work to report the event, or maybe even perceive personal negative outcomes in retaliation for reporting the information. In addition, "pencil whipping" may be occurring in relation to mandatory quota systems in safety reporting (Ludwig, 2014). Finally, in reality, most of the time the variables assessed using safety measures are typically safe. For these reasons, the variance from safety measurement is reduced to a small standard deviation.

Organizational components of readiness. In the beginning stages of analytics, the organization must first define the output desired through analysis as well as the mission and vision for the analysis project (Comuzzi & Patel, 2016; Eybers & Hattingh, 2017; Gao et al., 2015). Secondly, leadership must support the analytics process by providing personnel with the necessary capabilities to aggregate, clean, analyze, and visualize data (Comuzzi & Patel, 2016; Eybers & Hattingh, 2017; Gao et al., 2015) and drive the strategic value of the desired analysis. Technology must be acquitted and/or built in the form of platforms, software, databases and storage to produce quality data (Comuzzi & Patel, 2016; Eybers & Hattingh, 2017; Gao et al., 2015). Measurement culture, or the degree to which all levels of employees are willing to report behavioral and event incidents, must be mature in order to provide valid inputs to the database.

DART first examines data quality because data quality is essential to run any level of analysis (e.g. descriptive to prescriptive). Data quality, however, is affected by enabling organizational components also assessed in DART, including rules and operations, foundational infrastructure, and measurement culture.

Rules and operations. Rules and Operations refers to operations related to data collection, as well as rules of variable scaling and coverage. The readiness factor of Rules and Operations is assessed across three levels. First, in order to run analytics, the organization must have access to a range of variables that can be used to predict outcomes. Second, the data must be collected with common variables, so that un-centralized data can be connected. Third, the speed at which the data is updated is assessed.

Adequate coverage. Adequate coverage measures the extent to which key safety indicators in the measurement framework and variable matrix (e.g., pre-incident “leading” indicators, “lagging” outcome variables and process measures such as behavioral observations and inspections) are adequately covered within the data. Leading indicators are a measure of a company’s safety in the form of proactive measures that provide information about lower-impact incidents such as employee reports of close calls (e.g., near misses) and minor injuries. Process measures are created during safety activities and job processes to prevent risks and hazards from creating injuries. Examples of common process measures include safety audits, safety action item fulfillment, and behavioral observations. Lagging indicators are a measure of a company’s injury statistics, like the frequency and severity of OSHA recordable injuries and resulting lost time (i.e. time out of work), as well as company costs related to workers compensation, lost production, and insurance costs.

It is widely accepted that leading and lagging indicators are necessary to proactively target and improve occupational safety. A comprehensive safety management system should contain multiple metrics that monitor different dimensions of safety and performance (CCPS, 2009). Therefore, it is imperative that enough variables are adequately covered by organizations that wish to incorporate analytics into their safety programs.

Harmonization. Data harmonization refers to the availability of common data scaling and formatting across disparate variables and databases that allows the data to be linked and combined. For example, Microsoft Excel© uses the VLOOKUP formula to combine data from different sheets; the formula requires a common, unique variable by which the data sets can be matched. The same process is used within Big Data, which necessitates the availability of a common variable to match across datasets for combination into a single database. These most likely include demographics such as names, employee numbers, departments, dates, tasks, etc. These demographics can be as specific as employee names/ID number, can be aggregated into work teams or department, or be as general as a calendar unit (e.g. week, month, or quarter).

Velocity. Velocity refers to the frequency at which data enters a database and is updated (SAS, 2019). Examples of data velocity are the frequency of interval updates (e.g. daily, weekly, monthly, quarterly, or annually) of historical, batch, and real-time data feeds. Historical data, and data updated in larger intervals, is sufficient to describe what events have occurred in the past (e.g. descriptive and diagnostic analytics) and make predictions on what may occur in the future (e.g. predictive analytics) but in order to run prescriptive analyses, real-time data updates are essential, as prescriptive analytics has the capability of making automatic adjustments based on nuanced changes in the data stream. Within OHS, certain technologies such as wearable ergonomic smart belts with sensors can provide real-time information measuring pressure, ambient temperature, speed, and direction or angle of body movement. Most safety data, however, is collected through manual reports, either on paper or in an electronic system, and this delay between occurrence and report may prevent some organizations from being able to use those data sources for prescriptive analytics.

Foundational Infrastructure. Many times, organizations have multiple database technology platforms (i.e. foundational infrastructure) to help manage big data, and as a result, OHS departments may house their data differently than human resources, finance, or other functions. The database source has implications for analytics (e.g. how compatible platforms are for integration, whether raw or aggregated data is available to be extracted from the system, or what variables are chosen to be stored because of the original purpose of the data collection). To run analyses on variables contained in separate databases, it is essential to first merge data into one location. Statistical packages, such as R, can run analyses on separate databases, as long as there is a common variable within each database that the package can use to match variables with. In order to create these actionable databases to run analytics, conflicts, inconsistencies, or contradictions among data that is stored separately must be mitigated (Cai & Zhu, 2015). With small databases, it is easy to assess data quality via manual programming or searches through the values, but as this is impossible for large data, data quality must be assessed at each source.

As noted previously, safety data is typically collected to discover and mitigate specific issues after the measurement (e.g., inspection, audit, or submission of concerns) is conducted. Often the process of measurement is the lever used to mitigate these issues, where inspections lead to fixes and behavioral observation leads to feedback. Thus, the data is secondary to the measurement and is not managed afterward. Because of all these reasons, data can end up on paper forms or self-made spreadsheets in a variety of scaling, including nominal and text, and can be influenced by the reporting persons' biases. Accordingly, foundational infrastructure refers to the maturity of the organizational environment and technology processes devised to acquire, store, manage, and extract knowledge from data

(Comuzzi & Patel, 2016). Included in the DART factor of foundational infrastructure are (a) personnel with technical skills to manage and analyze data, and (b) centralization of the data.

*Personnel infrastructure***.** Personnel refers to the availability and expertise of key personnel necessary to carry out the technical processes of working with Big Data. Such expertise includes: (a) ensuring availability of data while minimizing cost (e.g. data management), (b) developing and maintaining predictive and forecasting models while establishing common analyses and reusable processes to reduce execution time and cost (e.g. analytics modelling), and (c) leadership oversight to define strategies and tactics that ensure relevance of analyses.

*Centralized database***.** Centralized database refers to the degree to which data are stored or can be readily combined into a central database. It will be rare for organizations to have one database for storing big data, unless the organization is utilizing advanced technologies such as data lakes for storing unstructured data or disparate databases, which can be restructured, aggregated, and transformed as later required (Quix, Hai, & Vatov, 2016).

Measurement culture. Law and Ruppert (2013) describe the collection of data as a social system, of which culture plays a large part. Data measurement culture impacts the entirety of the process through which data is collected. These processes include people interacting with the data entry forms in context of their work. Law and Ruppert (2013) thus posit that since these processes are heterogeneous arrangements between technology and humans, active social patterns emerge concerning data collection (e.g. employees conduct more observations and log reports at the end of a quota cycle). The authors further describe the communication of data within organizations as potentially political in their circulation (e.g. management purposefully may not discuss data findings with front-line workers). Therefore, measurement culture targets the willingness of employees to interact with date entry forms. This willingness may be affected by potential political ramifications (e.g., negative job outcomes), how data is presented to employees, or how improvements made based on the data are marketed (Beer, 2015). Employee participation may be a function of political assurances and feedback to employees showing improvements that have been made because of their participation.

Measurement culture refers to the extent to which employees and management are willing to provide valid accounts of what is happening in the workplace by completing inspection forms, conducting observations, or reporting close-call incidents. A company can have the best infrastructure possible, but the system will be ineffective for analysis and improvement if employees are not willing to participate in that system with integrity. Measurement culture assesses the willingness of employees to provide data voluntarily and regularly as well as their commitment to record information correctly, honestly, and in a timely manner. If data are not recorded at the time of the occurrence, the information is more subject to bias and some information may be lost, inaccurate, or missing due to faulty memory. Additionally, the culture surrounding reporting (e.g. the employee's comfort with speaking about incidents or risks, or protection from retaliatory practices) and management use of the data to find problems and make positive changes may impact the employee's willingness to report accurate information (Kagan & Barnoy, 2013). Two components of measurement culture are thus assessed: employee participation, and management concern for safety.

Employee participation. Employee participation in safety initiatives has been found to improve safety outcomes (Hagge, McGee, Matthews, & Alberle, 2017)**.** Within OHS, employees are a necessary component of data collection for hazard identifications, near miss identifications, observations and checklist completions, audits, and inspections. The extent to which employees participate in the process and report has an impact on both safety outcomes and quality of data. As previously mentioned, a factor that may impact data quality related to employee participation is "pencil whipping," which occurs when an employee fills out a report without an event taking place. Pencil whipping happens for many reasons, such as within mandatory quota systems, where an employee may feel pressured to fill out a certain number of reports regardless of whether a recordable event took place, or fear of reporting accurately due to negative repercussions from management (Ludwig, 2014).

*Management action***.** OSHA (2016) recommends that managers build safety culture by (a) encouraging employees to participate in the program, (b) encourage workers to report safety and health concerns, (c) involve workers in all aspects of a safety program, and (d) remove barriers to participation and reporting. Culture around reporting and data collection is additionally impacted by management concern (Frazier, Ludwig, Whitaker, & Roberts, 2013), which includes transparency about the purpose of reporting, support of time and cost of reporting in relation to production pressures, and use of the data collected by employees to make informed decisions to improve safety.

DART Preparation

Subject matter experts. The principle researcher, along with a team of graduate and undergraduate students, met frequently with teams from two participating companies, which included leadership (e.g. vice president), safety leadership (e.g. global health and safety director), safety coordinators, and additional representatives as needed (e.g. human resources generalists and information technology professionals). The research team from Appalachian State University set biweekly meetings with both organizations, to discuss matters related to the DART and the analytics projects used to compare against the DART findings.

The research team conducted on site meetings with each organization's subject matter experts to create a comprehensive list of hypothesized relationships within safety analytics and a list of the subsequent variables within those relationships. Appendix A outlines what occurred during that meeting, and summarizes the list of questions identified and the organization-specific metrics of interest. The research team SMEs distilled these resources from both organizations into a final Measurement Framework and Variable List. These tools are made available along with the DART assessment to guide data discovery and collection.

Measurement framework. Potential hypothetical relationships were documented by SMEs based on relationships found consistently in safety literature, in prior statistical analyses done by the SMEs, or suggested by trends identified in internal databases. Additional hypothetical relationships came from the SMEs' organizations' respective measurement dashboards whose spreadsheet algorithms and portrayal aggregations suggest hypothetical associations between data.

The Subject Matter Experts within each organization then diagramed their Measurement Framework as a visualization (i.e., a fishbone diagram) of the hypothesized relationships representing potential predictors, precursors to injury (e.g., first aid cases, close call reports), and additional mediator/moderator variables related to safety outcomes, including injuries. The Measurement Framework is a visual aid contained within a DART user manual to guide organizations to consider all of the possible relationships between organizational variables and safety outcomes (e.g. safety, operations, personnel, finance, maintenance, engineering, procurement). This diagram also provides a stimulus for organizations as they begin to collect information on what metrics and variables they have access to and wish to run analytics on. The Measurement Framework can be viewed in Appendix B.

Variable list. During the on-site meeting, the research team and SMEs from the host organizations brainstormed a list of the variables to be collected based on the Measurement Framework. The result was a list of the variables that commonly (i.e. as determined through literature and practitioner experience) lead to safety outcomes, as summarized within the Measurement Framework, including aspirational variables that the company is not currently measuring but promise to be related to injury reduction. The targeted variables were catalogued with a generic variable name. The resulting Variable List provides a listing of all the potential variables capable of contributing significant variance to analytic formulas. This list also provides a comparison to assess current metrics and direct the development of new metrics. More immediately, the variable list directs information gathering on the quality of current metrics (e.g. operational definition, scaling) and estimations of the availability of the data (e.g. location of data, how to retrieve) as the SMEs work through their self-assessment. The Variable List can be viewed in Appendix C.

DART Administration: Self-Assessment

The DART approach includes a systematic process of discovery and data collection prior to assessment of these factors (Figure 1). The Measurement Framework and Variable List are two visual representations of the organizational data that the DART factors are assessed against. Subject matter experts from the organization should view these two tools and begin collecting information on the variables in order to complete DART with the most possible accuracy. This information can be informal notes on the quality and format of the data, how the data are collected, and where the data are stored.

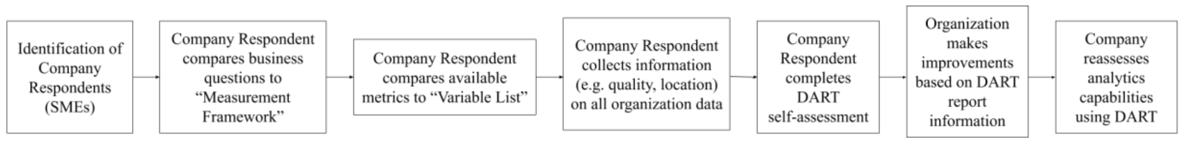


Figure 1. DART Data Collection and Discovery Process.

Once all necessary discovery information has been collected, the readiness of the organization for analytics is assessed against the organizational-level DART readiness factors. Each variable contained in the Variable List is then assessed against the Data Quality readiness factors in a self-assessment. Finally, DART provides scoring for readiness factors that will update as the organization improves. This scoring provides diagnostic information on the factors and the variables that lack analytic readiness.

The DART is designed to be a self-assessment tool. Each data and organizational component (Table 1) and the corresponding factors are self-rated using specific anchored criteria. Each self-rating is comprised of three stages: Low maturity, average maturity, and optimal maturity. These stages correspond to the CMM maturity stages, but are simplified to a three-point scale to reduce the time needed to complete the self-assessment by safety professionals (Preston & Colman, 1999). Each of these ratings have verbal criteria descriptions to aid the organization in this self-assessment of readiness. The verbal criteria guide the organizational representative as they assess their data against the DART readiness factors.

Table 1

Summary of DART Readiness Factors with their definitions

Success Factor	Readiness Factors	Definitions
Data Quality	Validity	Refers to the extent to which measures accurately represents the “real world” phenomenon targeted
	Reliability	Refers to consistency of measurement across time and units
	Variability	Refers to the ability of a measure to detect differences across time and units
Rules and Operations	Adequate Coverage	Measures the extent to which the things we want to look for in relation to safety outcomes are represented in data collection
	Velocity	Refers to the frequency with which data are collected, entered and updated in our databases
	Harmonization	Refers to having common demographics (e.g. who, what, where, when variables) across datasets that allow for data to be linked
Foundational Infrastructure	Personnel Infrastructure	Refers to the availability of key personnel with the necessary expertise to carry out technical processes of working with big data
	Centralized Database	Refers to the degree to which data variables are stored or can be readily combined into a central database
Measurement Culture	Employee Participation	The extent to which employees participate in the process and reporting of safety matters
	Management Concern	Support and encourage employees to participate, includes transparency about the purpose of reporting

The DART assessment Excel © workbook (Appendix D) contains four sheets. The first and second sheets are where respondents enter their ratings. For user ease, the organizational-level ratings are separated from the variable-level ratings. The organization-level ratings are assumed to remain consistent across individual metrics, as they refer to common resources or practices that each function, division, or organization would share. The respondent need only rate the organization-level factor once, rather than rating the factor for each variable. In comparison to this, each variable must be rated individual for quality. The second sheet has these variable-dependent ratings. The third sheet has a summary of all the ratings, and the respondent is able to make any adjustments to the ratings. The final sheet aggregates and scores all ratings.

Organizational level ratings. The first sheet of the DART Excel © workbook guides respondents to assess the organizational-level readiness factors: (a) Rules and Operations, covering Adequate Coverage, Harmonization, and Velocity, (b) Foundational Infrastructure, covering Personnel Infrastructure and Centralized Database, and (c) Measurement Culture, covering Employee Participation and Management Action.

Each organizational readiness factor has scaling criteria (see Table 2) to help raters determine maturity stages. These maturity stages are color coded within the Dart Excel © workbook. Green corresponds to optimal maturity, yellow corresponds to average maturity, and red corresponds to low maturity. The respondent will view each criteria definition and self-rate the stage of maturity for each readiness factor.

Table 2

Rating Criteria for Each Maturity Stage of the Organization-Level Readiness Factors

Success Factor	Readiness Factor	Maturity Stage	Criteria
Rules and Operations	Adequate Coverage	Low	The organization has data pertaining to less than a third of the Measurement Framework and Variable Matrix
		Average	The organization has data pertaining to more than half of the Measurement Framework and Variable Matrix identified variables
		Optimal	The organization has data pertaining to all of the Measurement Framework and Variable Matrix identified variables
	Harmonization	Low	Demographics are unique to the work function and cannot be linked to data from other areas
		Average	Demographics allow for data to be connected at the division/department/crew level
		Optimal	Demographics allow for data to be connected at individual or task level
	Velocity	Low	Data is collected, entered, and updated monthly
		Average	Data is collected, entered, and updated weekly or daily
		Optimal	Databases are updated as data is collected in real time
Foundational Infrastructure	Personnel Infrastructure	Low	The organization does not have employees with skills to manage large datasets and run analyses beyond Excel ©
		Average	The organization has personnel skilled in managing large data and running analyses but the personnel is limited to functions like IT or finance
		Optimal	Leadership engages in the formulation of business analytics questions, and has resourced an expert to work across functions
	Centralized Database	Low	Different functions have their own various databases, but nothing is centralized
		Average	Each function has their own centralized database

Success Factor	Readiness Factor	Maturity Stage	Criteria
		Optimal	Data across the organization is centralized, or the company uses technology that can perform a single query to all databases
Measurement Culture	Employee Participation	Low	Employees never participate in reporting
		Average	Employees sometimes report to supervisors or safety coordinators but do no direct reporting
		Optimal	Employees fill out reports on safety issues and participate actively in the safety process
	Management Concern	Low	Managers do not talk about safety or encourage participants to get involved
		Average	Managers talk about safety reports or encourage participation, but don't do both regularly
		Optimal	Managers talk about safety reports and encourage participation

A graphic representation then provides an estimate of the overall maturity of the organizational-level readiness factors. In this visualization, the organization can easily see which readiness factors are rated poorly, and which are optimal. The graphics provide a diagnosis for where the organization should invest in making organization-level improvements in order to optimize their analytics capabilities.

Variable-level ratings. As Data Quality can vary across metrics, each variable in the Variable List must be assessed individually. The respondent references the Variable List and inputs their organization-specific metrics into the second Excel © sheet. Respondents must rate each metric in the variable list for Data Quality (see Table 3) by rating validity (i.e. the extent to which measures accurately represents the “real world” phenomenon targeted), reliability (i.e. the consistency of measurement across time and units), and variability (i.e. ability of a measure to detect differences across time and units).

Table 3

Rating Criteria for Each Maturity Stage of the Variable-Level Readiness Factors

Success Factor	Readiness Factor	Maturity Stage	Criteria
Data Quality	Validity	Low	Data cannot be trusted (e.g. missing values or impossible values, like negative hours worked)
		Average	Data is generally trustworthy, (e.g. there are outliers of some occurrences that do not seem to be plausible)
		Optimal	The organization has validated that the data collected represents the phenomenon it purports to measure with no errors
	Reliability	Low	Data representing the same phenomenon are recorded differently and have different definitions
		Average	Data that represent the same phenomenon are recorded differently, despite having the same definition
		Optimal	Data representing the same phenomenon have the same definition and are recorded the same
	Variability	Low	Measures tend to get the same values over time
		Average	The measurement is sensitive enough to measure when something out of the ordinary happens
		Optimal	The measure is sensitive enough to detect the differences between everything commonly occurring

DART Scoring and Results

After the Data Quality has been assessed for each variable, the respondent is directed to turn to sheet 3 of the Excel © workbook. This sheet has a color-coded summary of all the maturity stage assessments. The table also translates the scale into summary statements: 0 = “Do Not Have,” 1 = “Needs Improvement,” and 2 = “Ready” (Appendix D). The SME is able to edit any of the organizational readiness success factor ratings that may vary for an individual metric. For example, an organization may have centralized data except for one department. The respondent has the opportunity to change the ratings for that one department to a red low maturity rating.

Aggregate scoring. The final sheet of the DART Excel © workbook aggregates all of the ratings into a score for overall readiness, which can be used to compare DART results over time, as well as a score for readiness at each of the four levels of analytics (i.e., descriptive, diagnostic, predictive, and prescriptive analytics). The self-ratings from the organization-level and variable level assessment are given a numeric value (i.e., Low Maturity = 0, Average Maturity = 1, and Optimal Maturity = 2) corresponding to the criteria within each of the rating scales (see criteria listed above).

Each level of analytics is defined by the readiness factors required. The research team provides initial factor weights and the minimum criteria for each level of analytics maturity using their expertise and experience in analytics and safety measurement. For example, data quality is foundational for all analyses which aim to deliver meaningful business outcomes (Jugulum, 2016). As such, average maturity of data quality is required across all levels of analytics, though optimal maturity is required for advanced analytics (i.e. predictive and prescriptive analytics) as the outcome of analyses is of increased actionable importance. This logic is expanded upon for prescriptive analytics; prescriptive analytics offer the highest value of output, and thus require the highest level of optimization (Frazzetto, Nielsen, Pedersen & Šikšnys, 2019). Additionally, prior statistical analyses have demonstrated some necessary components for each level of analytics (e.g. a prescriptive analytics study in manufacturing required optimal data acquisition, connectivity, data storage, data processing and control; Vater, Harscheidt, & Knoll, 2019). These initial estimates will be replaced by empirically validated coefficients as the DART factors are validated and loaded against actual analytic maturity levels related to clinically relevant effect sizes across a large number of organizations.

Each readiness factor is given a weight that represents how essential the factor is to analytics. Data Quality readiness factors are weighted highest (i.e. consistent with other readiness assessments; Venkatraman et al, 2016) at 15%, as clean and reliable data is essential to glean any reliable insight from analysis. Rules and Operations are weighted as second-most-important, along with Personnel Infrastructure, at 10%. Finally, Centralized Database and Measurement Culture are weighted at 5% each, as these factors will affect the ease with which analytics can be run or the quality of data as moderators, but are not base requirements for analytics.

Descriptive analytics require average maturity ratings of "1" across Validity, Reliability, Variability, and Adequate Coverage readiness factors (i.e. minimum score of 4). Diagnostic analytics additionally requires average maturity (1) across Velocity, Harmonization, and Personnel (i.e. minimum score of 7). Predictive analytics require average maturity (1) of Adequate Coverage, Velocity, and Centralized Database sub-factors, and optimal maturity (2) of the Validity, Reliability, Variability, Harmonization, and Personnel Infrastructure sub-factors (i.e. minimum score of 13). Prescriptive analytics require Validity, Reliability, Variability, Adequate Coverage, Harmonization, Velocity, Personnel Infrastructure, and Centralized Database to be optimized (2) for a minimum score of 16.

To calculate a readiness score for each analytics level, ratings are assessed against these minimums (see Table 4). If the minimum requirement is not met, the score is calculated as a zero and does not contribute to the ratio. For example, if the company did not meet the minimum requirement for velocity (2), for prescriptive analytics, then that score would revert to 0, even if it was average maturity. If all other requirements were met, the unweighted score would be 14. Each readiness factor score is then multiplied by the weight assigned. The organization's readiness score for each level of analytics is calculated by dividing the weighted total self-rating scores by the weighted required scores.

Table 4

Scoring System for Readiness Factors: Weights and Minimum Required Ratings to Achieve levels of Analytic Maturity

Readiness Factors	Weight	Descriptive	Diagnostic	Predictive	Prescriptive
Validity	0.15	1	1	2	2
Reliability	0.15	1	1	2	2
Variance	0.15	1	1	2	2
Adequate Coverage	0.10	1	1	1	2
Velocity	0.10		1	1	2
Harmonization	0.10		1	2	2
Personnel	0.10		1	2	2
Centralized Database	0.05			1	2
Employee Participation	0.05				
Management Action	0.05				

Note: Low maturity corresponds to a score of 0, average maturity corresponds to a score of 1, and optimal maturity corresponds to a score of 2. Average scores for each factor, across all variables, were compared to the minimum threshold of each analytics level and weighted.

The overall readiness score for the organization is also calculated from these numeric scores (Table 4). The numeric average ratings for each readiness factor are weighted and summed. This number is then divided by 2, the optimal score, for an aggregated total readiness score, which represents the degree of optimization.

The average readiness score does not correspond to any level of analytics, but reflects overall analytics process optimization, and serves as a diagnostic tool. Organizations can compare their overall readiness scores over time to check for improvement. Additionally, organizations can change their self-ratings for specific sub-factors to see how their readiness score is affected, thereby giving them a tool for assessing where to most effectively invest in improvements.

DART Recommendations

The final purpose of the DART is to provide information to a participating company on where to focus efforts in improving their data measurement systems in order to improve their analytic capabilities. This information spans the organizational and variable levels by looking at readiness factors and the individual variables.

In order to determine recommendations in a systematic way, criteria for recommendations is two-fold: the suboptimal readiness factors and variables are identified, and the optimal readiness factors are identified. Sub-optimal areas are defined as being >50% ready, while Optimal areas are over 75% ready. Each readiness factor is given a percent readiness score by taking the average rating (the score populated to the final DART Excel © workbook) and dividing the rating by the optimal score of 2. This represents the degree to which that readiness factor is approaching optimization. The variable quality is also assessed to determine which variables need attention within the current measurement system. The readiness for the variable quality is calculated by dividing the quality ratings (i.e. validity, reliability and variability) by 2, the optimal score. This provides a percent readiness score for each of the three Data Quality readiness factors across all 23 variables, allowing for specific metric recommendations.

After calculating the percent readiness for the readiness factors and the variable quality, the organization or external consultant can determine which areas are optimal and sub-optimal. Optimal areas (i.e. >75% ready) are current areas of strength within the company that can be further optimized with relatively low cost or effort, as optimization is within easy reach. Sub-optimal areas (i.e. >50% ready), on the other hand, may require extensive investment to improve, but these investments will ultimately lead to larger changes in analytics readiness.

By assessing these two categories, the DART is able to provide two courses of action for organizations to consider when attempting to improve their capabilities: areas that will constitute "quick wins" and gains, and areas which require forethought, planning, and investment, but which will be essential to improving analytics readiness.

Pilot Study

DART was created in reference to maturity models developed for use in other sectors, such as supply chain operations and healthcare, and seeks to improve analytical capabilities within safety by being the first model of data analytics maturity in OHS. Maturity models are typically evaluated by comparing them to existing solutions, seeking feedback from domain experts, and assessing validity against organizations (Pöppelbuß, & Röglinger, 2011). As previously stated, there has not been a published record of a maturity model assessment for the safety industry. Thus, DART was developed in partnership with domain experts and was piloted in two separate multinational organizations to explore the tool's validity.

The research methodology for the pilot (visually represented in Figure 2) included:

a) Subject matter experts were identified across each participating organization.

b) SMEs and the research team completed an Excel © data sheet that compiled information on all organization-specific metrics and outlined some business questions that could be answered with analytics (see Appendix E for example from Company B). This sheet was used as a reference as the respondents viewed the Variable List to compare the availability of metrics and assess specific variables.

c) Data was collected from each of the organizations over a period of a year.

d) The research teams conducted on-site documentation reviews, interviews, and focus group meetings gathering information regarding elements of the DART components.

e) DART assessments were completed by internal Subject Matter Experts and employees were administered the Measurement Culture Surveys (Appendix F) across the two organizations. Subject Matter Experts from the research team familiar with the host organization's data also completed the assessment to establish criterion validity.

f) Rudimentary analytics were run to explore which level of analytic capabilities the companies actually had.

g) A comparison was then made between the results from DART and the results from the conducted analytics.

h) The research team compiled DART results into a report of recommendations for each company.

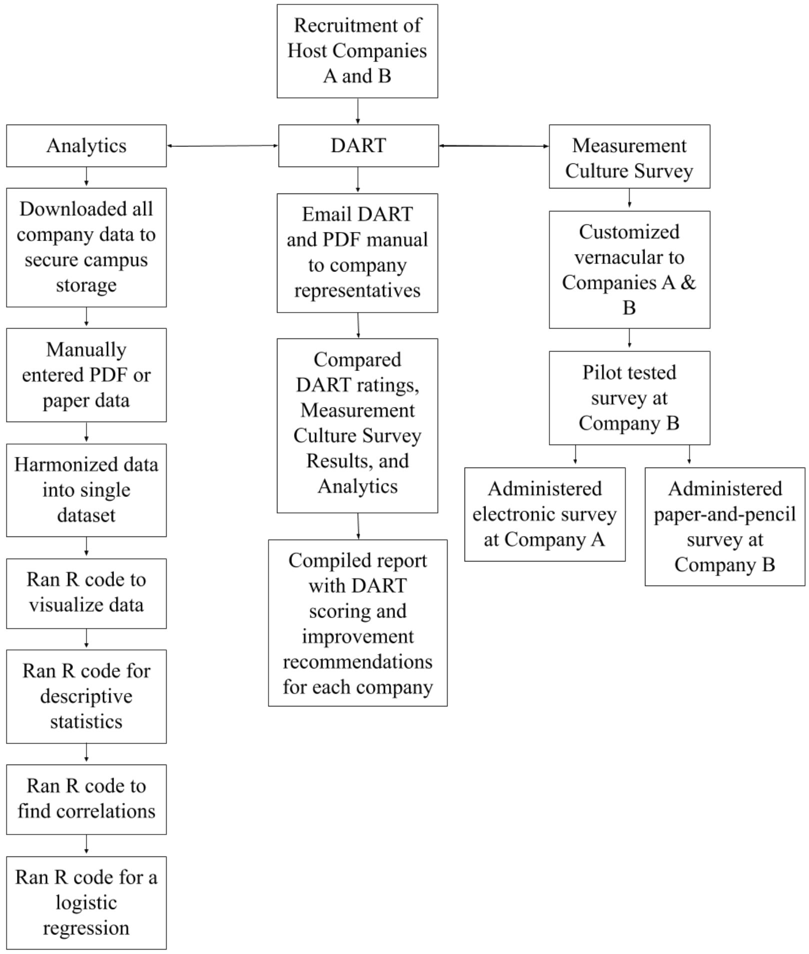


Figure 2. Research process.

Participating Organizations

Officers at each participating company and the research team at Appalachian State University have signed data sharing agreements and appropriate non-disclosure agreements. This research has been found exempt from IRB per Appalachian State University (20-0059), and full documentation can be found in Appendix G.

Company A. Company A is a global specialty chemical company producing a broad range of products for transportation, industrials, building and construction, and consumables. The organization employs approximately 15,000 people worldwide in more than 53 manufacturing locations. Company A is an American Chemistry Council Responsible Care company and its dedicated Health, Safety, Environment & Security (HSES) Management Systems have been reviewed by third parties for proper Responsible Care and or ISO (Europe) certification. Three large divisions (i.e., Technology, Manufacturing, and Maintenance) were used as initial samples. Company A used a vendor for its database management who provide data entry, storage, and reporting services, who agreed to provide access to the organization's non-aggregated data and its code for aggregation and reporting.

Company B. Company B is a parent company of five distinct textile companies and is the leading global brand that provides innovative solutions and premium textile products throughout the world. Company B services 25,000 customers in over 100 countries and across six continents via a global footprint that includes 37 facilities and over 15,000 employees. The company offers customers a breadth of textile products and solutions ranging from performance-driven and specialty apparel fabrics, high quality denims, advanced technical fabrics, and premium industrial and consumer sewing threads. The research team worked exclusively within the thread division of one textile company.

Data Collection. The research team began the process of collecting archived data from each company and moving it into a confidential filing system (i.e. Ustor) for cleaning and analysis. The research team used R to combine and clean datasets, and began testing the hypothesized relationships contained in the Measurement Framework. They catalogued the level of analytics they achieved for each analysis (i.e. descriptive, diagnostic, predictive, or prescriptive).

Measurement Culture Survey. The research team discussed the components of the measurement culture survey with each host organization, tailoring the language of the survey to vernacular used frequently in the organization. The appropriate power was determined for the sample of each organization. At Company A, the survey was released electronically. For Company B, the survey was administered on paper in groups so that non-native English-speaking participants could ask for verbal clarification. The survey was anonymous. For company B, participants who took the survey in groups turned them in to a safety coordinator, who turned them over to the research team without reviewing them.

Employee Participation. Items on the culture survey used to assess employee participation (willingness to participate in the data collection by completing forms, conducting observations, and reporting other safety issues) included:

- Reporting minor injuries (e.g. injuries requiring only first aid treatment and which do not involve medical treatment, loss of consciousness, restriction of work or motion, or transfer to another job; OSHA, 2019)
- Reporting near misses (e.g. close calls in which a worker might have been hurt if the circumstances had been slightly different; OSHA, 2001)
- Using safety forms to report safety information
- Supervisors reinforcing employee participation by responding quickly to solve safety problems

Management action. Items used to assess management action in the culture survey (e.g. the perceived use of safety data by managers and the perception of feedback provided after data is collected) included:

- Supervisor encourages employees to participate in decisions which affect safety
- Supervisor encourages employee involvement in audits, inspections, and behavior observations and perform these regularly
- Supervisors regularly ask employees about safety concerns and listens to our ideas

- Supervisor talks about things learned from incidents
- Supervisors use reports to make improvements
- Incidents that have the potential for serious injury (P-SIFs) are thoroughly investigated with accurate information.

DART Assessment

Organizational representatives (SMEs) identified for their expertise and engagement were provided the DART Excel © workbook and a PDF manual. The PDF manual (Appendix H) includes steps for preparing to take the DART assessment. The company respondents were given a week to complete the DART workbook, and were directed to ask questions of the research team when clarification was needed. No company respondents needed additional clarification.

Reliability Check

Two members of the research team also completed a DART assessment for each organization based on their knowledge of the organization's available metrics and data management. Research Team members had worked with Company A for eighteen months prior to completing the self-assessment. Another set of Research Team members worked with Company B for eight months prior to completing the assessment. Both teams of researchers had biweekly contact with host representatives and had visited the company's headquarters and manufacturing plants multiple times.

The research team worked with the host companies to fill out a Variable List, and gained knowledge of the coverage of organization-specific metrics as compared to that list. The student researchers downloaded all data from the company's host databases, and assessed steps needed to move all data into a singular database. This required the researchers to input and code data corresponding to each item in the variable matrixes separately. They then ran summary statistics to determine if the data contained outliers or errors. The researchers came into contact with the velocity of data updates as they updated data sources and aggregated data by day, week, or month for trends. The research team performed the necessary steps (e.g., manual data entry, VLOOKUPs, and data combining in R) to gain harmonization where possible, and were able to assess the maturity of harmonization of the data as it was when first handed over to the team.

The research team came into contact with internal personnel who represented the available expertise to collect measures, manage code, and run statistical analytics on organizational data. The team was also able to accurately assess the extent to which the data was centralized at each host company because they were required to download the data from each disparate database and combine the data in order to run analytics.

These researchers, with detailed knowledge and experience with the host organizations' data, personnel, and infrastructure served as reliability raters. Inter-rater agreement between company representatives and research team respondents was calculated for each readiness factor across 23 variables in the Variable List, the average readiness factor rating, and the DART aggregate readiness score.

Pilot Study Results

DART self-assessments were completed by one company representative and two research team respondents at Company A. Self-assessments were completed by three company representatives and two research team respondents at Company B.

Each company's DART readiness factor ratings per respondent are assessed, as well as the average ratings for the company respondents compared to research team respondents.

All scores are compared across raters to assess the reliability of the DART in measuring analytics maturity. Finally, the results from the Measurement Culture survey are used to further assess the reliability of the DART Measurement Culture self-assessment ratings.

These ratings provide diagnostic information for each company. A few recommendations for improving analytics optimization at each organization are provided.

Company A

DART Readiness Factor Ratings. The Company A Respondent largely rated their organization-level as optimally ready; only 4 readiness factors were rated as average maturity, and no readiness factor was rated at a low maturity stage. In stark contrast to this, the research team largely rated the company at an average maturity stage, with only four sub-factors reaching optimal maturity, across both raters, and one factor, harmonization, as being low maturity. Table 5 demonstrates the differences in scoring across each of the readiness factors that led to differences in aggregate scores.

Table 5

Summary of Company A's DART Average Readiness Factor Ratings by Respondent, Averaged Across all 23 Variable Ratings from the Variable List

Readiness Factors	CAR1	RT1	RT2
Data Quality			
Validity	2	2	1
Reliability	1	1	1
Variability	1	1	1
Rules and Operations			
Adequate Coverage	2	1	2
Velocity	2	1	1
Harmonization	1	0	1
Foundational Infrastructure			
Personnel Infrastructure	1	1	1
Centralized Database	2	1	1
Measurement Culture			
Employee Participation	2	2	2
Management Action	2	1	1

Note: Table displaying all sub-factor results from the Data Analytics Readiness Assessment. CAR denotes "Company A Respondent;" RT denotes "Research Team."

The average ratings of the research team compared to the corporate-level respondent demonstrates a similar trend (Table 6). Except for the sub-factors on which both sets of participants agreed (i.e. Reliability, Variability, Personnel Infrastructure, and Employee Participation), the Company A respondent rated each sub-factor higher than the research team.

Table 6

Comparison of Company A's DART Average Readiness Factor Ratings to the Research Team Ratings

Success Factors	CAR	Average Ratings RT
Data Quality		
Validity	2	1.5
Reliability	1	1
Variability	1	1
Rules and Operations		
Adequate Coverage	2	1.5
Velocity	2	1
Harmonization	1	0.5
Foundational Infrastructure		
Personnel Infrastructure	1	1
Centralized Database	2	1
Measurement Culture		
Employee Participation	2	2
Management Action	2	1
Total Aggregate Score (AVG)	**75**	**56.5**

Note: Table displaying average ratings from the Data Analytics Readiness Assessment. CAR denotes "Company A Respondent;" RT denotes "Research Team."

DART Readiness Scores. Company A's SME's ratings resulted in an overall score of 75% optimized (Table 7). The self-assessment determined that the organization is 100% prepared for descriptive and diagnostic analytics, 67% prepared for predictive analytics, and 50% ready for prescriptive analytics. In contrast, the research team respondents' ratings of Company A's organizational and variable level readiness factors resulted in an overall score of 55% and 58% optimized. While the research team respondents agreed on descriptive and diagnostic analytic capabilities, Research Team Respondent 1 rated Company A as 42% ready for predictive analytics and 13% ready for prescriptive analytics, and Research Team Respondent 2 found Company A to be 33% ready for predictive analytics and 13% ready for prescriptive analytics.

Table 7

Summary of DART Optimization Scores for Each Level of Analytics Per Respondents at Company A

Level	Respondent	Descriptive	Diagnostic	Predictive	Prescriptive	Overall
Company A	CAR 1	100%	100%	67%	50%	75%
	RT 1	100%	100%	42%	13%	55%
	RT 2	100%	100%	33%	13%	58%

Note: Table displaying all aggregate score results from the Data Analytics Readiness Assessment. CAR denotes "Company A Respondent;" RT denotes Research Team Respondent.

Reliability Analysis. Inter-rater agreement was calculated between raters, and averaged for an overall inter-rater agreement percentage. For each readiness factor, the score of the first respondent was compared to the second. The number of rating agreements across all 23 variables rated for that factor were added together. This number was then divided by 23 for a percent agreement. For example, Company Respondent 1 and Research Team respondent 1 agreed on Validity readiness factor ratings for thirteen variables. Thirteen divided by 23 variables gave an inter-rater agreement percentage of 57% for the Validity readiness factor. These percent agreements were then averaged across all comparisons for an average inter-rater agreement. These inter-rater agreement percentages are used to determine the reliability of the DART as a measure of analytics readiness when assessed for consistency across raters. Table 8 demonstrates the results of (a) comparing Company A Respondent 1 ratings to Research Team Respondent 1 ratings, (b) comparing Company A Respondent 1 ratings to Research Team Respondent 2 ratings, (c) comparing Research team respondent 1 ratings to Research Team Respondent 2 ratings, and (d) averaging the inter-rater agreement for Company A DART respondents.

Table 8

Inter-rater Agreement (IRA) for Readiness Factor Scores Across all 23 Variables in the Variable List.

Readiness Factors	IRA CR to RT1	IRA CR to RT2	IRA RT1 to RT2	Average IRA
Data Quality				
Validity	57%	61%	52%	57%
Reliability	57%	52%	52%	54%
Variability	57%	39%	30%	42%
Rules and Operations				
Adequate Coverage	0%	65%	35%	33%
Velocity	0%	0%	100%	33%
Harmonization	0%	100%	0%	33%
Foundational Infrastructure				
Personnel Infrastructure	100%	100%	100%	100%
Centralized Database	0%	0%	83%	28%
Measurement Culture				
Employee Participation	100%	70%	70%	80%
Management Action	0%	0%	100%	33%

Note: Table displaying inter-rater agreement of scores from the Data Analytics Readiness Assessment. Company A Respondent 1 was compared to Research Team Respondent 1; Company A Respondent 1 was compared to Research Team Respondent 2; Research Team Respondent 1 was compared to Research Team Respondent 2.

Inter-rater agreement across variable ratings for all of the readiness factors was low, except for Personnel Infrastructure (i.e., 100% inter-rater agreement). Readiness factors demonstrating higher reliability (i.e. >50% inter-rater agreement) are (in order of decreasing average agreement) Employee Participation, Validity, and Reliability. Readiness Factors demonstrating low reliability across variable ratings (i.e. <50%) are (in order of decreasing average agreement) Variability, Adequate Coverage, Velocity, Harmonization, Management Action, and Centralized Database.

The aggregate score is based on the average ratings for each readiness factor across all 23 variables in the Variable List. Therefore, the agreement between raters for the aggregate score was also assessed. The percent agreement across all participants rating Company A's corporate data maturity ranged from 67%-100% (see Table 9), with 100% agreement on the sub-factors (a) reliability, (b) variability, (c) personnel infrastructure, and (d) employee participation. Across the research team self-ratings, there was higher agreement; there was 100% agreement on seven sub-factors, but three sub-factors were rated differently.

Table 9

Agreement Across Average Readiness Scores for all DART respondents versus the Research Team

Readiness Factors	Percent Agreement Across All Raters	Percent Agreement Across Research Team
Data Quality		
Validity	67%	50%
Reliability	100%	100%
Variability	100%	100%
Rules and Operations		
Adequate Coverage	67%	50%
Velocity	67%	100%
Harmonization	67%	50%
Foundational Infrastructure		
Personnel Infrastructure	100%	100%
Centralized Database	67%	100%
Measurement Culture		
Employee Participation	100%	100%
Management Action	67%	100%

Note: Table displaying agreement of ratings from the Data Analytics Readiness Assessment.

Company B

DART Readiness Factor Ratings. The summary of the individual ratings across all of Company B respondents outlines the trends in self-ratings (Table 10). The success factor of Data Quality had almost complete agreement, with only one respondent rating the Reliability sub-factor higher. The same is true for the success factor of Foundational Infrastructure; all respondents rated all sub-factors as low maturity, except for one sub-factor rating by one respondent. The scores across the other success factors and sub-factors are also similar, though there is some variance.

Table 10

Summary of DART Results Company B

Success Factors	CBR1	CBR2	CBR3	RT3	RT4
Data Quality					
Validity	1	1	1	1	1
Reliability	1	1	2	1	1
Variability	1	1	1	1	1
Rules and Operations					
Adequate Coverage	1	1	1	1	1
Velocity	0	1	1	0	0
Harmonization	0	1	1	0	0
Foundational Infrastructure					
Personnel Infrastructure	0	1	0	0	0
Centralized Database	0	0	0	0	0
Measurement Culture					
Employee Participation	1	0	1	1	2
Management Action	1	1	1	1	2

Note: Table displaying all sub-factor results from the Data Analytics Readiness Assessment. CBR denotes "Company B Respondent;" RT denotes "Research Team."

A comparison on Company B's DART ratings from company respondents to the research team demonstrates similar findings to the Company A comparison. On average, company B respondents rated the organization's maturity higher than the research team respondents, with the exception of the success factor Measurement Culture (Table 11). The aggregate total readiness scores across both sets of respondents are more similar for Company B than for Company A.

Table 11

Comparison of Company B's DART Average Readiness Factor Ratings to the Research Team Ratings

Readiness Factors	Average Ratings CBR	Average Ratings RT
Data Quality		
Validity	1	1
Reliability	1.33	1
Variability	1	1
Rules and Operations		
Adequate Coverage	1	1
Velocity	0.67	0
Harmonization	0.67	0
Foundational Infrastructure		
Personnel Infrastructure	0.33	0
Centralized Database	0	0
Measurement Culture		
Employee Participation	0.67	1.5
Management Action	1	1.5
Total Aggregate Score (AVG)	**42.67**	**35.5**

Note: Table displaying average ratings from the Data Analytics Readiness Assessment. CBR denotes "Company B Respondent;" RT denotes "Research Team."

DART Readiness Scores. The research team respondents had 86% agreement in their aggregated total readiness scores (e.g., calculated by dividing the difference in the two percent scores by the average of the two scores, subtracting this from 1, and multiplying by 100), while the organization self-rated at a higher percent overall and had 59% agreement across the highest to lowest total aggregate readiness scores. Three of the raters' final aggregate scores were in the 30% range (i.e. 33, 33, and 38), while Company B Respondent 2 rated the company as 45% optimized, and the final Company B Respondent rated the company as 50% optimized.

Company B Respondent 1 and the two research team respondents had complete agreement in the percent readiness for the different levels of analytics (i.e. 100% ready for descriptive analytics, 57% ready for diagnostic analytics, 8% ready for predictive analytics, and 0% ready for prescriptive analytics) though their final total aggregate readiness scores differed. Company B Respondent 2 rated the company as 100% ready for both descriptive and diagnostic analytics, 17% ready for predictive analytics, and 0% ready for prescriptive analytics. The final Company B respondent rated the company as 100% ready for descriptive and diagnostic analytics, 33% ready for predictive analytics, and 13% ready for prescriptive analytics.

Table 12

Summary of DART Optimization Scores for Each Level of Analytics Per Respondent at Company B

Level	Respondent	Descriptive	Diagnostic	Predictive	Prescriptive	Overall
Company B	CBR 1	100%	57%	8%	0%	33%
	CBR 2	100%	100%	17%	0%	45%
	CBR 3	100%	100%	33%	13%	50%
	RT 3	100%	57%	8%	0%	33%
	RT 4	100%	57%	8%	0%	38%

Note: Table displaying all results from the Data Analytics Readiness Assessment. CAR denotes "Company A Respondent;" CBR denotes "Company B Respondent;" RT denotes Research Team.

Reliability Analysis. Inter-rater agreement was also assessed across the Company B raters. As Company B had three company respondents, their inter-rater agreement was calculated first, along with an average agreement percentage (Table 13). Company Respondent 1 was found to have ratings most consistent with their peers, and was thus chosen as a comparator for the research team respondents. Table 13 demonstrates the results of (a) comparing Company A Respondent 1 ratings to Research Team Respondent 1 ratings, (b) comparing Company A Respondent 1 ratings to Research Team Respondent 2 ratings, and (c) comparing Research team respondent 1 ratings to Research Team Respondent 2 ratings.

Table 13

Inter-rater Agreement (IRA) for Readiness Factor Scores Across all 23 Variables in the Variable List for each Company B Respondent

Readiness Factors	IRA CR1 to CR2	IRA CR1 to CR3	IRA CR2 to CR3	Average IRA
Data Quality				
Validity	35%	43%	39%	57%
Reliability	35%	83%	39%	54%
Variability	48%	70%	48%	42%
Rules and Operations				
Adequate Coverage	26%	100%	26%	51%
Velocity	48%	0%	35%	28%
Harmonization	35%	0%	17%	33%
Foundational Infrastructure				
Personnel Infrastructure	35%	100%	35%	57%
Centralized Database	100%	100%	100%	100%
Measurement Culture				
Employee Participation	17%	100%	17%	45%
Management Action	4%	100%	4%	36%

Note: Table displaying inter-rater agreement of scores from the Data Analytics Readiness Assessment.

As with Company A, inter-rater agreement for all across variable ratings for the readiness factors was low, except for one (i.e., Ratings for Centralized Database had 100% average inter-rater agreement). Readiness factors with higher reliability (i.e. inter-rater agreement >50%), in order of decreasing average agreement, were Validity, Personnel Infrastructure, Reliability, and Adequate Coverage. Readiness factors with lower reliability (i.e. inter-rater agreement <50%), in order of decreasing average agreement, were Employee Participation, Variability, Management Action, Harmonization, and Velocity.

Inter-rater agreement between Company A Respondent 1 and the Research Team was much higher (Table 14). These respondents gained 100% average inter-rater agreement in four readiness factors: Velocity, Harmonization, Personnel Infrastructure, and Centralized Database. Four additional readiness factors (e.g., Adequate Coverage, Reliability, Variability, and Validity) had average inter-rater agreement ≥50%. Only two readiness factors (e.g. Employee Participation and Management Action) had average inter-rater agreement below 50%.

Table 14

Agreement Across DART Ratings from Company B Corporate Respondent 1 to Research Team Respondents

Readiness Factors	IRA CR1 to RT1	IRA CR1 to RT2	IRA RT1 to RT2	Average IRA
Data Quality				
Validity	57%	35%	57%	50%
Reliability	48%	48%	65%	54%
Variability	61%	39%	61%	54%
Rules and Operations				
Adequate Coverage	65%	61%	78%	68%
Velocity	100%	100%	100%	100%
Harmonization	100%	100%	100%	100%
Foundational Infrastructure				
Personnel Infrastructure	100%	100%	100%	100%
Centralized Database	100%	100%	100%	100%
Measurement Culture				
Employee Participation	74%	0%	0%	25%
Management Action	39%	0%	35%	25%

Note: Table displaying inter-rater agreement of scores from the Data Analytics Readiness Assessment.

The analysis of respondent agreement across the average score for each readiness factor performed for Company A was repeated for Company B (Table 15). For the corporate level, there were three respondents for Company B, and so agreement was assessed across (a) all respondents, (b) the Company B respondents, and (c) the research team respondents.

Table 15

Agreement Across Average Readiness Scores for all DART respondents versus the Company Respondents versus the Research Team

Success Factors	Percent Agreement Across Raters	Percent Agreement Company B Raters	Percent Agreement Research Team Raters
Data Quality			
Validity	100%	100%	100%
Reliability	80%	67%	100%
Variability	100%	100%	100%
Rules and Operations			
Adequate Coverage	100%	100%	100%
Velocity	60%	67%	100%
Harmonization	60%	67%	100%
Foundational Infrastructure			
Personnel Infrastructure	80%	67%	100%
Centralized Database	80%	67%	100%
Measurement Culture			
Employee Participation	60%	67%	50%
Management Action	80%	100%	50%

Note: Table displaying agreement of ratings from the Data Analytics Readiness Assessment Readiness Factors.

Across all raters, the highest agreement (i.e. ≥ 80%) was in the Data Quality and Foundational Infrastructure success factors and the adequate coverage and management action sub-factors. The three Company Respondents had a minimum agreement of two out of three respondents, and had 100% agreement on the subfactors (a) validity, (b) variability, (c) adequate coverage, and (d) management action. The two Research Team respondents had 100% agreement across all success factors except for Measurement Culture.

Culture Survey Results

DART Measurement Culture ratings were on a three-point Likert scale from 0-2. This scale was converted to a five-point Likert scale in order to compare to the survey scale more easily. A zero corresponds to a one, a one corresponds to a three, and a two corresponds to the highest rating, a five. The measurement culture survey results were averaged across two factors. Factor 1 corresponds to the readiness factor Employee Participation, and is represented by questions 1-5. Factor 2 corresponds to the readiness factor Management Action and is represented by questions 6-14.

The culture survey results were compared to the self-assessment of the Measurement Culture readiness factor in DART. This comparison shows rating inflation in self-assessment scores. Across all ratees for both companies, the self-assessment ratings were higher than the rating averages for the survey (see Table 16).

Table 16

Comparison Between DART Culture Self-Assessment and Survey Results

	DART EP	converted	Survey EP	DART MA	converted	Survey MA
CAR 1	2	5	3.83	2	5	4.19
RT 1	2	5		1	3	
RT 2	2	5		1	3	
CBR 1	1	3	3.87	1	3	3.82
CBR 2	1	3		2	5	
CBR 3	1	3		1	3	
RT 3	1	3		2	5	
RT 4	2	5		2	5	

Note: For comparability to the survey responses, the DART scores are converted to a 5-point scale, where a 0 corresponds to 1, 1 corresponds to 3, and 2 corresponds to 5. EP refers to employee participation, and MA refers to management action. Survey ratings are averages across all respondents (Company A n total = 348, Company B total n = 152).

Company A Respondent 1 and Company A research team respondents rated the sub-factor of Employee Participation as optimal (5), despite the survey results of 3.83. Company A Respondent 1 also rated Management Action optimally (5), while the survey results were 4.19. The Company A research team respondents rated Management Action at a 3, and were closer to the division-level survey results. Further summary statistics (e.g., mean, median, standard deviation, skewness) for each question on the survey can be found in Appendix I.

All Company B respondents and one of the research team respondents rated Employee Participation at average maturity (3), while one research team respondent rated the readiness factor optimally (5). These ratings corresponded well to the front-line employee-answered survey result of 3.87. For Management Action, two Company B respondents rated the sub-factor at average maturity (3) while one Company B respondent and both research team respondents rated the sub-factor optimally (5). The average across these ratings is 4.2, which also corresponds well to the survey results of 3.82.

Action Step Recommendations

The final report back to each company provided scores and ratings for all factors and measurement culture survey results. In addition, for each factor that was sub-optimal (i.e. Readiness Factor Score below 50% optimized), the research team provided recommendations on improvements. These reports can be found in Appendix J.

Company A. In order to maximize predictive analytics, Company A should (a) ensure the availability of common variables that can be used to connect data, (b) dedicate personnel that can clean, aggregate, and analyze safety data, and (c) improve the sensitivity of safety measurement systems to allow them to capture more variance. Additionally, while the company has strong Employee Participation, it needs to improve their Management Action readiness factor. This can be done by communicating the purpose of data collection and increasing transparency for how data is used to improve safety.

Areas of strength (i.e. >75% ready) were also identified for Company A. Company A was strong in Adequate Coverage and had metrics on many of the essential predictors of safety outcomes. Company A should continue to improve the data quality of these measures to enhance the value of predictive analytics. Company A was also strong in their Employee

Participation readiness. Employee participation in reporting will affect data quality. Company A should continue to support and encourage employees to accurately report safety incidents in order to improve data quality of those measures.

Company B. The next level that Company B will be capable of achieving is diagnostic analytics. In order to maximize diagnostic analytics, Company B must improve their readiness in the Centralized Database and Personnel Infrastructure readiness factors.

In order to maximize safety analytics, Company B must create a centralized database where data variables can be stored or readily combined, and must dedicate manpower and additional resources to cleaning, aggregating, and running analyses on safety variables.

The DART also assessed Company B’s areas of strength. Company B had high readiness in factors Management Action and Reliability. This assessment recommends that Company B continue to show employees how data is used to make positive improvements in the organization, which will relate to improvements to the reliability and accuracy of employee reporting.

Analytics Maturity

The research team compared the DART scores to levels of analytics that they were able to achieve at each organization in order to assess the criterion validity of the tool. Each company gave the Appalachian State University Research Team access to their safety data, and the team cleaned, aggregated, and conducted the most advanced analytics they could.

Company A. The research team attempted to run analytical models at the descriptive, diagnostic, predictive, and prescriptive levels.

Descriptive analytics. Descriptive analytics consists of summary statistics and data visualizations. For each of the available variables, the research team was able to determine sums, means, and standard deviations across the measures. They were able to explore trends using this information, as the example in Figure 3 demonstrates. The team analyzed the difference in cumulative behavioral observations between when an incident occurs versus no incidents. It was determined that Company A had met the required maturity to run descriptive analytics.

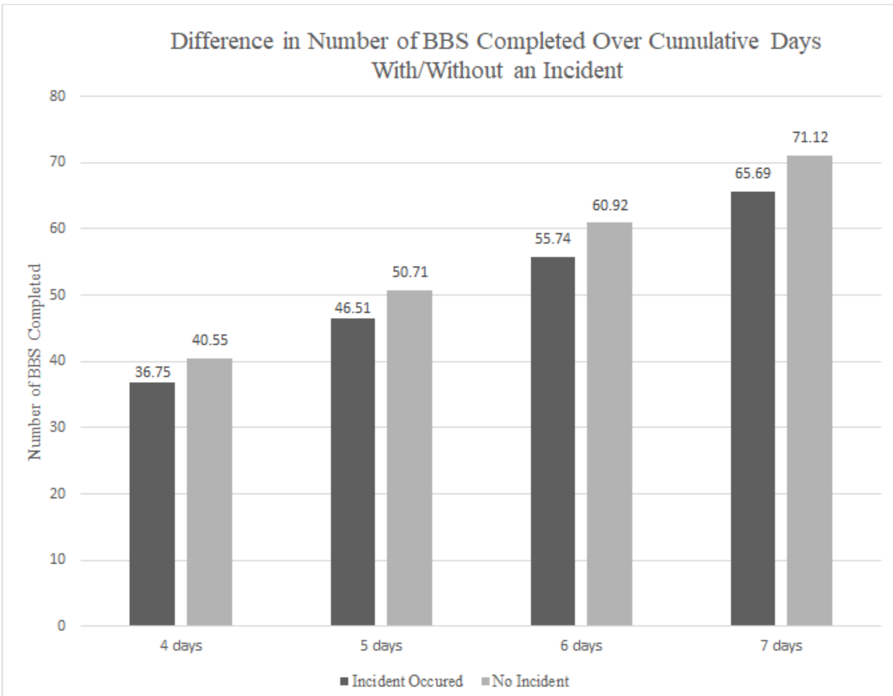


Figure 3. Graph Displaying the Cumulative Monthly Accident Rate by Fiscal Year and Plant

Diagnostic analytics. Diagnostic analytics begin to explore relationships between variables. The research team looked at various correlations in order to assess whether variables had relationships with incidents. Table 17 is representative of one of these analyses. This diagnostic analysis was assessing whether there was a relationship between product changes or line changes and (a) incidents, (b), number of inspections, or (c) number of behavioral observations. There was not a significant relationship between the number of daily product or line changes and incidents, but there was an increase in incidents on days that had a line change (M = .02, SD = .14), compared to days without a change (M = .00, SD = .06), t =

2.10, $p < .05$. Furthermore, the correlation between product changes and inspections was significant ($r = .11$, $p < .01$), indicating that as the number of product changes increased, employees performed more frequent inspections. From these analyses, it was determined that Company A had met the required maturity to run diagnostic analytics.

Table 17

Means, Standard Deviations, and Correlation of Company A Variables

	1	2	3	4	5
1.Product Changes	---				
2.Line Changes	.62***	---			
3.Observations	.06†	.07†	---		
4.Inspections	.11**	.06	.25***	---	
5.Incidents	.02	.04	.02	.03	---
Mean	3.82	1.22	3.40	.39	.01
SD	2.92	1.43	3.08	.72	.11

† $p < .10$; * $p < .05$; ** $p < .01$; ***p < .001

Predictive analytics. The research team was also able to complete predictive analytics reports for the company by running logistic regressions. The predictive power of a logistic regression is calculated by using a c-value statistic. The c-value, measured on a 0.5 to 1 scale, indicates whether the model predicts random outcomes (e.g., scores closer to 0.5) versus the model perfectly discriminating the outcome (e.g., scores closer to 1). Task safety audits predicted incidents when lagged five (c=0.541), six (c=0.548), and seven days (c=0.542), and behavioral observations were predictive of incidents occurring the same day (c=0.561). The research team also predicted the likelihood of an incident occurring through a combination of predictors (e.g., crew, year, day or night shift, consecutive work days, whether a shift was near a holiday, and average wind speed; c=.714) and the likelihood of near misses occurring due to crew, year, day or night shift, whether a shift was near a holiday, and consecutive work days (c=0.65). These predictive reports had small predictive power, likely due to low variance in incident rates.

Prescriptive analytics. The research team was not able to conduct prescriptive analytics. After strong predictive relationships have been found and the velocity of data collection is increased, variables from Company A can be put into a model that will assess risk and suggest action in order to mitigate that risk.

Predicted vs. actual analytics maturity. The value of the output from each level of analytics suggested that Company A is capable of running descriptive and diagnostic analytics, however they are not optimized for predictive analytics. Improvement to their outcome variables' sensitivity to differences in incident severity (e.g., increased variability) would greatly improve their ability to run predictive analytics. This is supporting evidence that the DART scores, averaged across all respondents in this pilot (i.e., 100% ready for descriptive analytics, 100% ready for diagnostic analytics, 47% ready for predictive analytics, and 25% ready for prescriptive analytics), were able to accurately assess Company A's analytics readiness. The company respondent, however, had a higher self-assessment of readiness (e.g., 67% optimized for predictive analytics) than the research team respondents (e.g., 42% and 33% optimized for predictive analytics, respectively), and the research team's assessment are likely more accurate because they were conducting analyses.

Company B. The research team attempted to run analytical models at the descriptive, diagnostic, predictive, and prescriptive levels. The research team was given access to Company B's data, but it required extensive data entry and cleaning, as the majority of it was contained in unstandardized Excel © templates or scanned paper-and-pencil documents.

Descriptive analytics. The team was able to compile some descriptive statistics and visuals for the company. Figure 4 is an example of the descriptive analytics that were conducted on Company B's data. The research team created a visual using 15 years of data of total monthly accidents to represent the average monthly accidents occurring at each plant each year. These statistics and visuals provided value to the company, as they were able to prioritize which plants to focus further analyses. It was thus determined that Company B had the required maturity to run descriptive analytics.

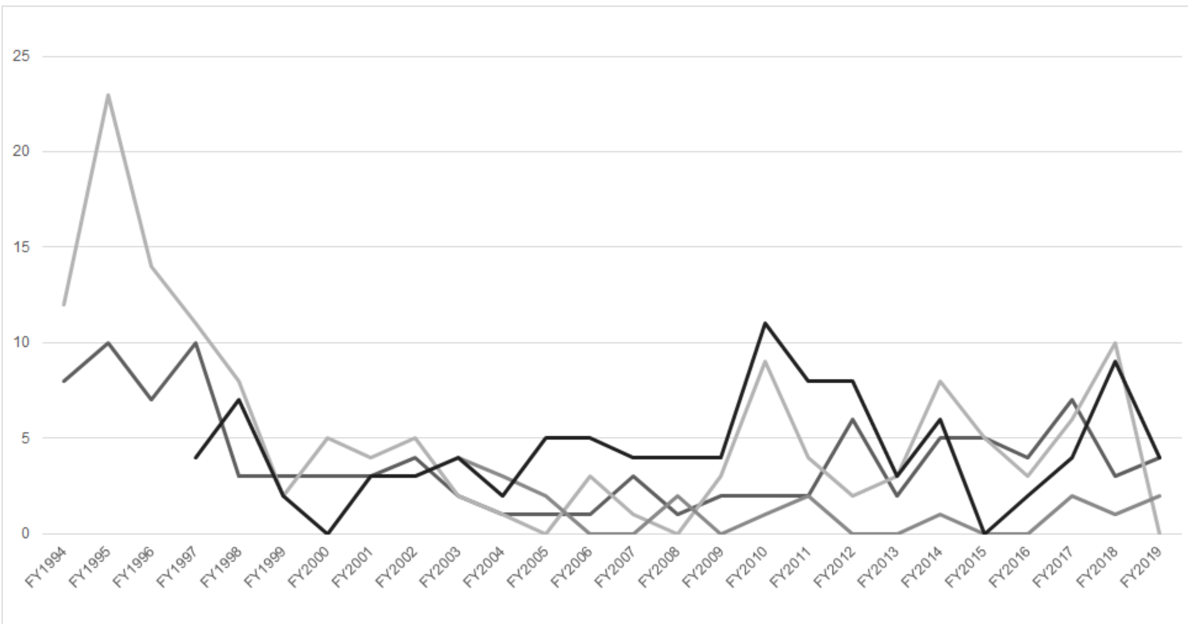


Figure 4. Average number of monthly accidents each fiscal year for each plant. Each line represents a different plant at Company B; the vertical axis represents an average of monthly accidents per fiscal year.

Diagnostic analytics. Due to the extensive data cleaning that was required at Company B, there was not adequate coverage of the required variables to attempt to determine relationships or run correlations in the fiscal quarter the DART was administered.

It was determined that Company B did not have the required maturity in order to run diagnostic analytics.

Predictive analytics. Predictive analytics was similarly limited by the availability of variables with which to run analyses. The research team was unable to run any predictive models, such as regressions, and it was thus determined that Company B did not have the required maturity to run predictive analytics.

Prescriptive analytics. As the most advanced level of analytics, prescriptive models cannot be run without the availability of adequate variables, frequent velocity of data updates, and predictive relationships. As these readiness factors were lacking, the research team was unable to run prescriptive analytics at Company B, and it was determined that the company did not have the required maturity.

Predicted vs. actual analytics maturity. The value of the output from each level of analytics suggested that Company B is capable of running descriptive analytics, but is not ready for diagnostic, predictive, or prescriptive analytics. This is additional supporting evidence that the DART scores, averaged across all respondents (i.e., 100% optimized for descriptive analytics, 74% optimized for diagnostic analytics, 15% optimized for predictive analytics, and 3% optimized for prescriptive analytics), were able to accurately assess Company B's analytics readiness. Two of the company respondents from Company B had inflated self-assessment scores (i.e. self-assessment determined 100% optimization for diagnostic analytics), while the other three raters had unanimous agreement, yet self-rated Company B's readiness lower, suggesting that the DART accuracy may depend on rater tendency or bias.

Discussion

The Data Analytics Readiness Tool is the first maturity model designed to assess analytic capabilities in the safety industry. It is also the first readiness assessment to further discern readiness at each of the four levels of analytics (e.g. descriptive, diagnostic, predictive, and prescriptive). Though the DART was built in reference to existing readiness assessments, the novelty of its focus required heavy reliance on subject matter experts in both data science and safety. In order to assess the reliability of the DART as a measurement tool of analytic capabilities, the tool was piloted at two multinational manufacturing organizations. The DART self-assessments completed by company representatives were compared to DART self-assessments completed by the research team in order to further quantify criterion validity.

The results of the DART self-assessment were found to be fairly accurate to the actual analytic capabilities within the two participating organizations in our case study. Preliminary validity analysis showed Company A to be fully capable of running analytics at the diagnostic level. While research students were able to run predictive models at Company A, c-values showed low predictive power. This corroborates the aggregate readiness score (averaged across respondents) of 47% ready for predictive analytics. The aggregate readiness scores for Company B (also averaged across respondents) also accurately assessed analytic capability. Company B was determined to be 100% ready for descriptive analytics, while only 74% ready for diagnostic analytics, as the lack of centralized data prevented any diagnostic analyses.

For each company, DART was able to determine specific recommendations for improvement at the variable and organization levels. The DART was built as a tool that can be revisited periodically. For example, companies can complete the DART annually and reference to assess the improvements that their investments have made on their analytic capabilities.

Reliability

There was a difference in the readiness factor ratings of the individual variables, suggesting deficits in reliability across DART users. Despite a higher level of agreement between the raters of each specific company, only three readiness factors were scored with >50% agreement across raters for both companies: Validity, Reliability, and Personnel Infrastructure. Four readiness factors, Variability, Harmonization, Velocity, and Management Action, did not reach 50% agreement at either pilot. The maturity criteria may need to be revisited for these items in order to increase the accuracy and reliability of the measure. It is unclear at this moment whether the low reliability across raters is due to rater bias or error, or due to scaling. Preston and Colman (1999) found that three-point scales have the lowest test-retest reliability, but no studies have assessed the reliability of the number of scaling items for a behaviorally anchored rating scale (BARS), like the maturity criteria descriptions. It is important to note, however, that though rater agreement differed for each of the readiness factors at the variable-level, the DART proved to be robust against these differences, as the average readiness scores and DART aggregate scores were comparable across raters.

In addition to the rating differences across individual DART users, there was a rating difference between the company representatives and the research team respondents. For each pilot self-assessment, company representatives had more inflated ratings than research team respondents. This is likely due to familiarity with the data; while the company representatives have knowledge of company culture and metrics, they have less expertise in the types of analytics and what is necessary. The research team on the other hand, has expertise in data science, and spent time working with the company data, cleaning it, aggregating it, and running analyses, which ultimately lent itself to a more realistic picture of many of the readiness factors. In the future, it is likely best to recommend that DART is completed by an impartial external rater who has data analytics expertise and experience working with the company's datasets. If DART must be completed by a company representative, some frame-of-reference training may be required in order to more fully indoctrinate the representative to the necessary components of the maturity criteria.

The measurement culture survey was found to capture more variance in response that the DART self-assessment. This is to be expected, as the 2 sub-factors that it represents are covered across 14 questions, and the scaling is a five-point behaviorally anchored rating scale rather than a three-point Likert scale. Thus, even if it is not possible to run the survey across a fully-representative sample at the host organization, self-raters should use the Measurement Culture Survey to inform their ratings on that DART success factor assessment when possible, instead of relying only on the maturity stage criteria.

Limitations

The DART was compared to analytics results from two organizations. The goal to establish criterion validity was truncated. When data was collected and analytics was attempted, Company B showed to be at an immature level of readiness (i.e. optimized for descriptive analytics) and Company A was shown to be at a moderate level of readiness (i.e. optimized for diagnostic analytics). We had no examples of companies whose readiness allowed us to demonstrate higher maturity analytics (e.g., predictive and prescriptive analytics). The DART assessment needs to be validated across many different organizations that represent all levels of readiness and actual analytics maturity. While the two organizations assessed in this study were from manufacturing (albeit differing products), future research should validate the DART across a multitude of safety environments and industries (e.g. construction or roadway safety).

Another limitation to the current study is the differences in ratings between internal organization SMEs and the RT raters. For example, at Company B, the research team found that the DART recommended focusing on the descriptive level of analytics, while some company representatives found their DART assessments put them at 100% ready for descriptive and diagnostic analytics, with a higher percent readiness for predictive analytics. One respondent even rated the company as having some amount of optimization leading to a 13% readiness for prescriptive analytics. A major problem of the DART scoring across both organizations was this self-rating inflation whereby most company-based raters were shown to be more lenient in their assessment of capabilities compared to external objective raters. Further investigation is needed to determine if this difference in ratings was due to data science expertise and exposure to actual data, or if there were other causes. The DART manual and assessment needs to be developed further to gain rating reliability across users. The manual may need to include some frame-of-reference training to acclimate users to the maturity criteria system, or have more formalized steps prior to the completion of the DART workbook where the user gains exposure to additional exposure to the data by (a) reviewing each form of data or database and (b) answering questions on quality and availability. Additionally, though the maturity stage criteria were simplified for the user experience (i.e. the CMM five maturity stages were reduced to three), expanding the DART criteria may capture necessary variance across company readiness and may contribute to more accurate ratings.

Future Research

The Measurement Framework and Variable List used as the standard in the pilot study was created based on current literature and organization understanding of hypothetical relationships in among their metrics. The DART will likely adapt and change as literature and participating organizations discover additional relationships between variables and important safety outcomes.

Further research should be done to refine and validate the weighting and scoring system used in the DART. The self-ratings are reliant on the interpretation of the respondent, which is a common issue with maturity model scaling. This can be addressed by validating the scoring across many different respondents, and adjusting the weights as needed to combat common rater errors or biases. After validating DART across a range of organizations that have varied capabilities, it will be possible to statistically validate which sub-factors drive analytics capabilities (i.e. the required minimum ratings and weighted scoring system). Further research should also analyze the moderating effect of Measurement Culture on data analytics readiness. Measurement Culture added numeric value to the total aggregate readiness score, as it impacts optimization, but was not required for any level of analytics. Measurement Culture should be assessed to determine how it impacts other success factors, such as Data Quality, and this in turn may impact how Measurement Culture should be weighted and scored in DART.

Future research should statistically validate the results of the DART with actual attempts at analytic levels of maturity. Instead of comparing the DART ratings to visuals or predictive models using subsets of data, researchers could use all available data from the host organization to create a full model. While the research team was only able to run one type of predictive model (e.g. logistic regression) at Company A, a full picture of the available data will determine which analytics model would provide the greatest insight and which statistics would be necessary to evaluate the predictive power of the model. This evaluation may include assessing effect sizes, r^2, or p-values to determine the success of a company achieving their desired analytics maturity level. Further administrations of the DART with actual analytics runs will help to build empirically-derived weights and maturity level cutoffs.

The DART should provide an accurate picture of the value of the analytics output to drive improved decision-making. The purpose of DART is to provide a tool to organizations to assess and improve their capabilities so that they can utilize data to make more informed decisions. For example, in our pilot, we were able to determine that Company A, while able to run predictive analytics, should focus on diagnostic analytics, because finding relationships between variables will provide them with the strongest statistical output to determine where to focus safety initiatives. The predictive models had low statistical power; therefore, while the output is information that can be factored into decision-making, it should not be entirely relied on, as the predictor variables only explain a small part of the variance in the outcome (i.e. risk of incident). In other words, the entire DART system needs to be validated across many organizations, but the level-specific readiness scores also need to be validated against the strength of the output from the levels of analytics at those companies.

Implications

With safety analytics being at such a nascent stage, it is imperative that organizations assess their current capabilities and making improvements to their safety systems in order to drive down injuries and fatalities. The DART is one such tool that provides not only a realistic picture of current capabilities (e.g. guidance on what level on analytics the company is ready for) but also a diagnosis for where improvements may make a difference (e.g. low maturity success factor scores). An industry specific readiness assessment will give managers and safety professionals the ability to gain analytical insight, ultimately leading to improved, data-driven initiatives that will target and reduce risk.

References

Arnold, K. E., Lonn, S., & Pistilli, M. D. (2014). An exercise in institutional reflection: The learning analytics readiness instrument (LARI). *ACM International Conference Proceeding Series*, 163–167. doi: 10.1145/2567574.2567621

Ayhan, S., Costas, P., & Samet, H. (2018). Prescriptive analytics system for long-range aircraft conflict detection and resolution. *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, 239–248. doi: 10.1145/3274895.3274947

Barling, J., Loughlin, C., & Kelloway, E. (2002). Development and test of a model linking safety-specific transformational leadership and occupational safety. *Journal of Applied Psychology*, 87, 488-496.

Beer, D. (2015). Productive measures: Culture and measurement in the context of everyday neoliberalism. *Big Data & Society*, *2*(1), 205395171557895. doi: 10.1177/2053951715578951

Bureau of Labor Statistics [BLS]. (2018a). *Employer-reported workplace injuries and illnesses - 2017, USDL-18-1788.*

Bureau of Labor Statistics [BLS]. (2018b). *National census of fatal occupational injuries in 2017, USDL-18-1978.*

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, *14*, 1–10. doi: 10.5334/dsj-2015-002

CCPS. (2009). *Guidelines for Process Safety Metrics*. John Wiley & Sons.

Chi, C. F., Lin, S. Z., & Dewi, R. S. (2014). Graphical fault tree analysis for fatal falls in the construction industry. *Accident Analysis and Prevention*, *72*, 359–369. doi: 10.1016/j.aap.2014.07.019

Comuzzi, M., & Patel, A. (2016). How organisations leverage: Big Data: A maturity model. *Industrial Management and Data Systems*, *116*, 1468–1492. doi: 10.1108/IMDS-12-2015-0495

Cosic, R., Shanks, G., & Maynard, S. (2012). *Towards a business analytics capability maturity model*.

Eybers, S., & Hattingh, M. J. (2017). Critical success factor categories for big data: A preliminary analysis of the current academic landscape. *2017 IST-Africa Week Conference, IST-Africa 2017*, 1–11. doi: 10.23919/ISTAFRICA.2017.8102327

Fogarty, G. J. (2004). The role of organizational and individual variables in aircraft maintenance performance. *International Journal of Applied Aviation Studies*, 4, 73-90.

Frazier, C. B., Ludwig, T. D., Whitaker, B., & Roberts, D. S. (2013). A hierarchical factor analysis of a safety culture survey. *Journal of Safety Research*, *45*, 15–28. doi: 10.1016/j.jsr.2012.10.015

Fred, M. O., & Kinange, U. M. (2016). Creating value of human resource through analytics. *Asia Pacific Journal of Research*. Retrieved from www.apjor.com

Gao, J., Koronios, A., & Selle, S. (2015). Towards a process view on critical success factors in Big Data analytics projects. *2015 Americas Conference on Information Systems, AMCIS 2015*, 1–14.

Grossman, R. L. (2018). A framework for evaluating the analytic maturity of an organization. *International Journal of Information Management*, *38*, 45–51. doi: 10.1016/j.ijinfomgt.2017.08.005

Guo, S. Y., Ding, L. Y., Luo, H. B., & Jiang, X. Y. (2016). A Big-Data-based platform of workers' behavior: Observations from the field. *Accident Analysis and Prevention, 93*, 299–309. doi: 10.1016/j.aap.2015.09.024

Hadaya, P., & Pellerin, R. (2010). Determinants of construction companies' use of web-based interorganizational information systems. *Supply Chain Management, 15*, 371–384. doi: 10.1108/13598541011068305

Hagge, M., McGee, H., Matthews, G., & Aberle, S. (2017). Behavior-based safety in a coal mine: The relationship between observations, participation, and injuries over a 14-year period. *Journal of Organizational Behavior Management, 37*, 107–118. doi: 10.1080/01608061.2016.1236058

Huang, L., Wu, C., Wang, B., & Ouyang, Q. (2018). Big-data-driven safety decision-making: A conceptual framework and its influencing factors. *Safety Science*. doi: 10.1016/j.ssci.2018.05.012

Jugulum, R. (2016). Importance of data quality for analytics. In *Quality in the 21st Century: Perspectives from ASQ Feigenbaum Medal Winners* (pp. 23–31). doi: 10.1007/978-3-319-21332-3_2

Kagan, I., & Barnoy, S. (2013). Organizational safety culture and medical error reporting by Israeli nurses. *Journal of Nursing Scholarship, 4*, 273–280. doi: 10.1111/jnu.12026

Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers, 19*, 267–283. doi: 10.1007/s10796-016-9686-2

Law, J., & Ruppert, E. (2013). The social life of methods: Devices. *Journal of Cultural Economy, 6*, 229–240. doi: 10.1080/17530350.2013.812042

Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, *50*, 57-70. doi: 10.1016/j.ijinfomgt.2019.04.003

Lingard, H., Hallowell, M., Salas, R., & Pirzadeh, P. (2017). Leading or lagging? Temporal analysis of safety indicators on a large infrastructure construction project. *Safety Science*, *91*, 206-220. doi: 10.1016/j.ssci.2016.08.020

Ludwig, T. (2014). The anatomy of pencil whipping. *Professional Safety*, *59*, 47–50.

Mistikoglu, G., Gerek, I. H., Erdis, E., Mumtaz Usmen, P. E., Cakan, H., & Kazan, E. E. (2015). Decision tree analysis of construction fall accidents involving roofers. *Expert Systems with Applications*, *42*, 2256–2263. doi: 10.1016/j.eswa.2014.10.009

Mousanif, H. Saba, H., Douiji, Y., & Sayad, Y. O. (2014). From big data to big projects: A step-by-step roadmap. International Conference on Future Internet of Things and Cloud.

Nemati, H., & Udiavar, A. (2013). SCAX: Measuring Organizational Readiness to Embrace Supply Chain Analytics. *International Journal of Business Intelligence Research (IJBIR)*, *4*, 19-38. doi:10.4018/jbir.2010071701

Occupational Safety and Health Administration [OSHA] (2001). *Occupational Injury and Illness Recording and Reporting Requirements*. Retrieved from https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_id=16312&p_table=FEDERAL_REGISTER

Occupational Safety and Health Administration [OSHA] (January, 2012a). *Injury and illness prevention programs*. Retrieved from https://www.osha.gov/dsg/InjuryIllnessPreventionProgramsWhitePaper.html

Occupational Safety and Health Administration [OSHA] (2012b). *Predictive Analytics in Workplace Safety: Four "Safety Truths" that Reduce Workplace Injuries*. *A Predictive Solutions White Paper*. Retrieved from http://www.osha.gov/SLTC/etools/safetyhealth/mod1_estimating_costs.html

Occupational Safety and Health Administration [OSHA] (October, 2016). *Recommended Practices for Safety and Health Programs*. Retrieved from

Occupational Safety and Health Administration [OSHA] (2019). *Accident investigation*. Retrieved from https://www.osha.gov/dcsp/products/topics/incidentinvestigation/

https://www.osha.gov/shpguidelines/docs/OSHA_SHP_Recommended_Practices.pdf

Pfleeger, S. L. (1995). Maturity, models, and goals: How to build a metrics plan. *The Journal of Systems and Software*, *41*, 143-155. doi: 10.1016/0164-1212(94)00094-4

Poh, C. Q. X., Ubeynarayana, C. U., & Goh, Y. M. (2018). Safety leading indicators for construction sites: A machine learning approach. *Automation in Construction*, *93*, 375–386. doi: 10.1016/j.autcon.2018.03.022

Pöppelbuß, J., & Röglinger, M. (2011). What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management. *In 19th European Conference on Information Systems, ECIS 2011*.

Preston, C. C., & Colman, A. M. (1999). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*, 1–15. Retrieved from www.elsevier.com/locate/actpsy

Quix, C., Hai, R., & Vatov, I. (2016). Metadata extraction and management in data lakes with GEMMS. *Complex Systems Informatics and Modeling Quarterly*, *9*, 67–83. doi: 10.7250/csimq.2016-9.04

Ringen, K., Dong, X. S., Goldenhar, L. M., & Cain, C. T. (2018). Construction safety and health in the USA: Lessons from a decade of turmoil. *Annals of Work Exposures and Health*, *62*, S25–S33. doi: 10.1093/annweh/wxy069

Salas, R., & Hallowell, M. (2016). Predictive Validity of Safety Leading Indicators: Empirical Assessment in the Oil and Gas Sector. *Journal of Construction Engineering and Management*, *142*(10). doi: 10.1061/(ASCE)CO.1943-7862.0001167

Schultz, G. (2012). Don't investigate safety incidents... Predict and prevent them! *EHS Today, 5*, 59–60.

Snyder, R. A., & Fields, W. L. (2006). Measuring hospital readiness for information technology (IT) innovation: a multisite study of the organizational information technology innovation readiness scale. *Journal of Nursing Measurement*, *14*, 45-55. doi:10.1891/106137406780954324

Spruit, M., & Pietzka, K. (2015). MD3M: The master data management maturity model. *Computers in Human Behavior*, *51*, 1068–1076. doi: 10.1016/j.chb.2014.09.030

Stewart, D. (2013). A look at safety analytics. 'Retrieved from http://www.canadianminingjournal.com/features/a-look-at-safety-analytics/

Vallmuur, K., Marucci-Wellman, H. R., Taylor, J. A., Lehto, M., Corns, H. L., & Smith, G. S. (2016). Harnessing information from injury narratives in the 'big data' era: Understanding and applying machine learning for injury surveillance. *Injury Prevention, 22*, i34–i42. doi: 10.1136/injuryprev-2015-041813

Venkatraman, S., Sundarraj, R. P., & Mukherjee, A. (2016). Prototype design of a healthcare-analytics pre-adoption readiness assessment (HAPRA) instrument. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture*

Notes in Bioinformatics): *Vol. 9661 LNCS* (pp. 158–174). doi: 10.1007/978-3-319-39294-3_11

Appendix A
<u>Initial Meeting with Host Organization</u>

[redacted]-ASU Analytics Kick-off Meeting Recap

Thank you to everyone that could attend the [redacted]-ASU kick-off meeting on [redacted]! We had a productive meeting thanks to everyone's participation and involvement in brainstorming [redacted]'s business questions that could be answered with analytics. To review what we came up with and provide a recap for those unable to attend, this report will lay out the process and next steps for our analytics project.

Meeting Process

Safety coordinators, plant managers, and other [redacted] leadership came together with ASU faculty [redacted] and several students on the Safety Analytics Team to discuss what questions about safety at [redacted] that could be answered by running analytics. Our process to identify analytics possibilities at [redacted] consisted of these three steps: Business Question Brainstorm, Categorization of Questions, and Predictor Variable Identification.

Business Question Brainstorm

Each member of the [redacted] team was asked to write down some questions about safety that they believed would be beneficial to answer. Any and all questions were welcome, regardless of whether the data existed or was able to be answered through analytics, to get an idea of the types of questions [redacted] was interested in. All of the questions that came from this session can be found in Appendix A of this report.

Categorization of Questions

From the themes of the questions that were discussed, we came up with four categories that the questions could fit into. The analytics category became the specific area of focus for the project and thus was broken down further into sub-categories. The breakdown of questions are as follows:

- Safety Culture
- Training
- Human Resources
- Analytics
  - Personnel
  - Environment/Equipment
  - Leading Indicators
  - Workloads & Production

Predictor Variable Identification

To answer the questions in the analytics category, the teams brainstormed predictor variables associated with the questions in each analytics sub-category. The variables were put into a data matrix that lists the predictor variables in each row, and the columns explain how each variable is defined, measured, where the data is located, and who at [redacted] has access to the data. The list of these predictor variables can be found in Appendix B. A complete list of the variables and data matrix can be found as an attachment in the original email this report was included in.

Next Steps

Outcome Variables
Our next steps in this process is to identify the outcome variables of the questions in the analytics category and put them in the data matrix. For example, one of the outcomes discussed was OSHA recordables, so we need to know where that data is located, who has access, and a description of that data.

Prioritization of Analytics Questions
Secondly, we need to know which analytics questions are the most important to answer first for [redacted]. Questions prioritized first should have answers that would be of most value to the organization, and are possible to answer based on the data available.

[redacted]/ASU Update Meetings
We would like to set up bi-weekly update meetings to make sure everyone is on the same page with progress on the projects and an opportunity to troubleshoot any barriers on a regular basis. During these meetings we will come up with a timeline for project progress for the rest of the year.

Points of Contact
During the meeting, we discussed where data is located and who has access to it. Our understanding of the points of contact are listed below. If they are correct, we will need their contact information or a different point of contact.

- [redacted]: HR and Payroll Data
- [redacted]: Data for Days Since Last Lost Time
- Med Techs in HR: Hearing Quality and Metric Tests Data, Waste Produced Data
- [redacted]: Inventory Levels (WIP, Raw Materials, Finished Goods)
- Safety Coordinators per Plant: Individual Plant Data
- [redacted]: IT Data
- [redacted]: Leading Indicators Data
- [redacted]: Engineering Data
- Plant Control Person: Room Conditions Data

ASU Student Requirements
Determine if there is a non-disclosure agreement that the ASU students need to sign before looking at [redacted] data. If so, please let us know as soon as possible so we can get those forms signed and sent back to [redacted].

Safety Culture Questions

- How can we improve buy-in for safety culture?
- How can we get employees to speak up about safety issues?
- How do we measure improvement of safety culture?
- How can we make the safety culture tool sustainable?
- What variables have the most impact on safety culture?
- How can we differentiate between safety in the job vs. safety practices of the individuals?
- How do we reinforce safety behaviors of employees and decrease unsafe behaviors?
- What activities are contributing to lessened safety?
- Ask employees, "Would you allow your child to be employed here? Why or why not?"
- Will our changes to safety actually improve our company's safety?
- How can we balance safety and productivity?

Training Questions

- How can we verify that training is effective?
- How effective is safety training?
- How can we determine the effectiveness of a procedure in preventing accidents?
- How can we improve data literacy of employees?
- Are audits effective?

Human Resources Questions

- What are the expectations of what is and what isn't safe?
- How can we promote autonomy?
- How are our personnel's styles of learning and our methods of instruction lining up?

Analytics Questions

Personnel

- What is the correlation between safety culture and buy-in?
- What is the threshold for the number of hours worked before accidents start occurring?
- How many days in a row can people work before compromising safety?
- What traits are linked to individuals being involved in accidents?
- What behaviors improve safety?

Environment/Equipment

- What percentage of machines are working?
- How are hazards visually marked?
- Are machines working efficiently?

Leading Indicators

- Where and when will the next accident take place?
- What type of accident will take place next?
- Which leading indicators have the most impact on safety?

Workloads & Production

- How do product and quality issues affect safety?

    - Are we overloading our personnel?

Personnel

- Work Shifts
    - Number of hours
    - Days worked
    - Weekends worked
    - Overnight shifts
    - Team or individual work
    - Vacations
    - FMLA
    - Absences (sick days)
    - Sick on the job
    - Days worked in a row
- Job Type
    - Hourly/Salary
    - Production worker (incentive pay)
- Worker Characteristics
    - Tenure
    - Age
    - Male/Female
    - Height/Weight
    - Distance from Work
    - Demographics
- Organizational Commitment
    - Number of committees an employee is on
- Supervision
    - Change in Supervision
    - Tenure of Supervisor
    - Training Supervisors
    - Number of Direct Reports (manager to employee ratio)
- Disciplinary Actions
- Employee Training Activities
    - Safety
    - Trainings
    - Equipment
    - In-process and process changing
- Days and Hours Since Last Lost Time Event
- Hearing and Quality Metric Tests

Equipment and Environment

- Overhauling
- New Machinery
- Safety Work Orders
- Noise
- Machine of condition (age)

- Speed of Machine
- Defective Tools
- 5 S (standard of tools)
- Facility Condition
- Machine Changeover (modified machine)
- Number of History of Lap-ups
- Type of Product Run
- Machine Fixes
- Floor Conditions
- Threat Lubricant
- Safety Features (or lack thereof)
- Warning Lights (safety indicators)
- Labeling
- Spacing and Work Space Congestion
- Storage Equipment
- Mechanical versus manual
- Basic Guarding
- Single versus Multiple Operator
- Mats and Slip Mats
- Job Aids (e.g., needs to climb a ladder)
- Room Conditions

Leading Indicator Activities

- PPE Use
- Hazard ID
- Near Misses
- First Aid
- SNAP Cards
- SOC (Safety Observation and Coaching) Cards
- Zone Audits
- Safety Committee Meetings
- R3's
- One Point Kaizen
- Shift Meetings
- Safety Talks
- Accident Analysis
- Leading Indicator Scorecard
- Pre-shift Exercise
- Drills
- Detail, Lose, Run Report
- Insurance Audits

Workloads and Production

- Month of End
- Plant Production (pounds out the door)
    - Per Plant and Team

- Day of the Week (Week v Weekend)
- Reward
- Quality Inspections
    - Waste
- Inventory Levels
- Stockouts
- Weights (Handling of Materials)
- Heat (other conditions)
- Ergonomics
- Travel Distance/Steps
- Steps in the Process (info overload)
- Product Changeovers
- Customer Specific Requirements
- Departmental Touch-points

APPENDIX B

Measurement Framework

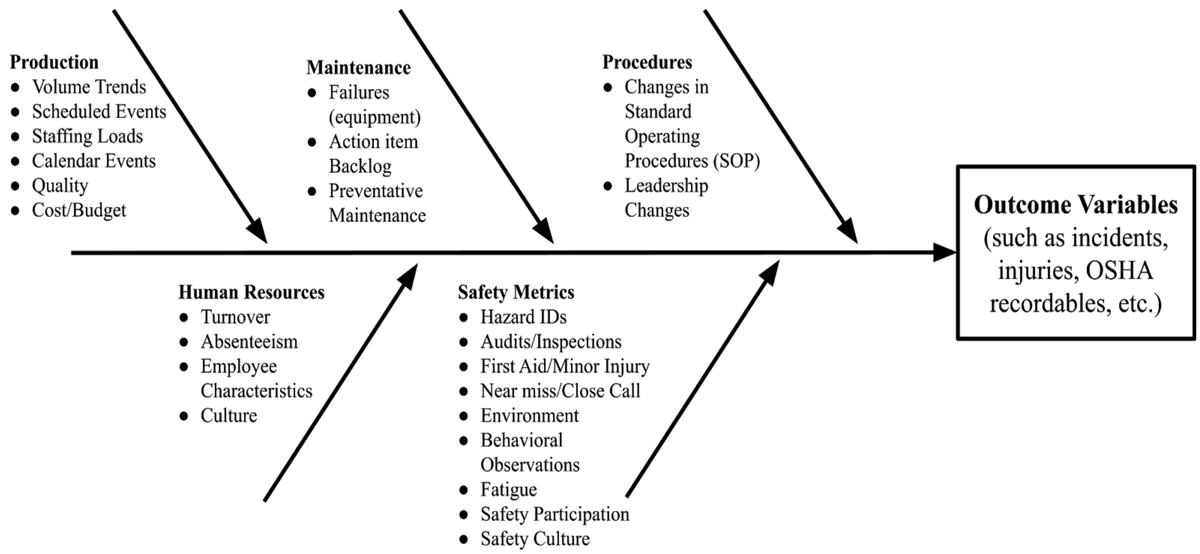


Figure 5. Measurement Framework visual describing hypothesized relationships between variables from different functions impacting outcome variables such as incidents.

APPENDIX C
Variable List

Table 18

Variable List consisting of the 23 important variables that may impact safety outcomes, along with their definitions.

Variable	Definition
Production	
Volume Trends	The quantity of product output at a given time.
Scheduled Events	These metrics relate to changes to the production that managers may make, such as switching a machine from one product to another.
Staffing Loads	These metrics relate to accounting for the amount of people working at a given time, and how much those employees are working.
Calendar Events	These are events that may center around a date on a calendar, like a holiday or vacation.
Quality	A measure of excellence or a state of being free from defects, deficiencies and significant variations. These metrics may also measure the opposite, such as errors or reworks.
Cost/Budget	These metrics would measure estimated costs, revenues, and resources over a specified period.
Maintenance	
Failures (Equipment)	These metrics track what machines have failed, how often, or why.
Action Item Backlog	These metrics measure unfinished tasks that need to be completed.
Preventative Maintenance	Data centered around equipment and facilities by tracking systematic inspection, detection, and correction of incipient failures either before they occur or before they develop into major defects.
Procedures	
Change Management	These metrics could track what occurs after a change in procedure, product, or could track leadership changes over time. These organizational metrics may cross-functions and help to provide a strategic lens.

Variable	Definition
Human Resources	
Turnover	A measurement of the number of employees who leave an organization during a specified time period.
Absenteeism	The measure of the number of employees that are absent from their scheduled shift.
Employee Characteristics	Demographics of the workforce that is being measured (e.g., tenure, sex, age, education).
Culture	The measure of the cultural norms of the workforce, including what they prioritize, talk about, and behave.
Safety Metrics	
Hazard ID	The metrics surrounding the reports made by employees of hazardous environments that could lead to safety incidents.
Audits/Inspections	These metrics track the frequency of audits and inspections of workplaces and workgroups to measure if the work is being done safely and if the environment is safe.
First Aid/Minor Injury Reporting	These metrics track minor injuries, such as small cuts, trips, or falls that may need first aid.
Near Miss/Close Call Reporting	This metric measures the number of instances where employees report that a safety incident did not occur, but almost did.
Environmental	Environmental metrics track adverse conditions such as weather, heat, wind, storms, etc.
Behavioral Observations	Metrics collected on behavioral observations may include video or checklists, either paper and pencil forms or electronic.
Fatigue	Measures tracking the point at which employees are likely to have a safety incident due to strain and fatigue caused by the nature of their work/workload.
Safety Participation	Measures of safety participation track whether front line employees are engaged with the safety process, including shift discussions, reporting, or investigating incidents.

Variable	Definition
Safety Culture	Safety culture measures the extent to which employees and managers think about, talk about, and are committed to having a safe workplace.

APPENDIX D

DART Excel © Workbook

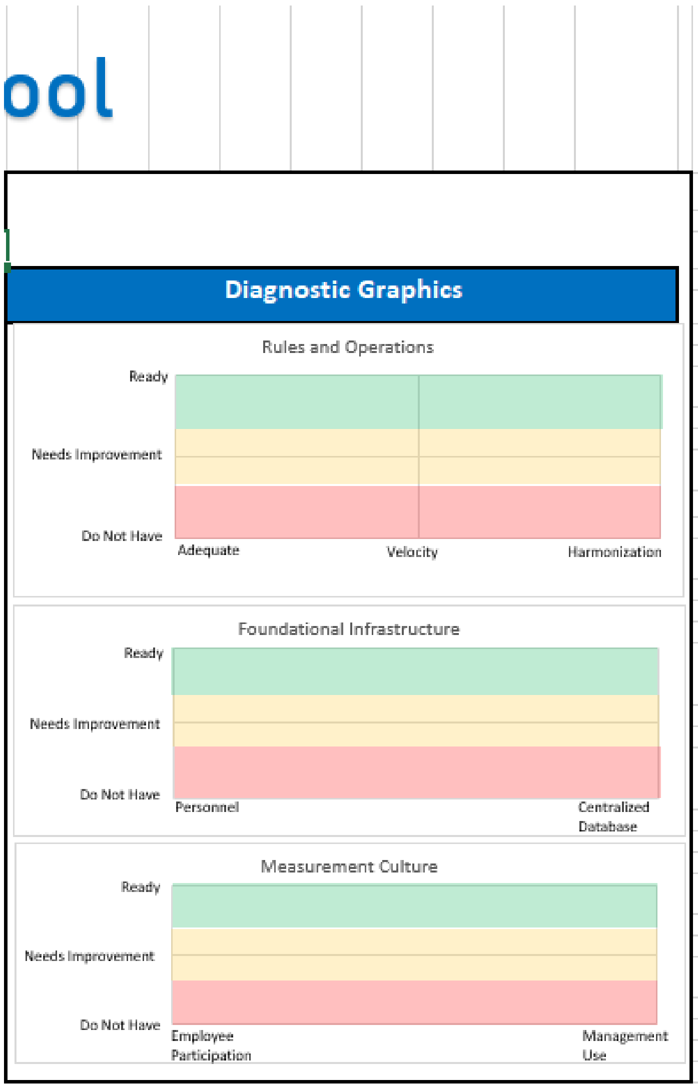
Data Analytics Readiness T

Organization-Level Ratings

Please identify the level of which you are completing the DART assessment for:

Rate the level's data readiness for the following organizational success factors:

Rating Scale:	Do Not Have	Needs Improvement	Ready
Rules and Operations	Rules and Operations refers to operations related to data collection, as well as rules of variable scaling and coverage		
Adequate Coverage			
Measures the extent to which the things we want to look for in relation to safety outcomes are represented	My organization collects data on about a third of the items in the variable list ☐	My organization collects data on about half of the items on the variable list ☐	My organization collects data on most of the items listed, as well as additional things ☐
Velocity			
Refers to the frequency with which data is collected, entered and updated in our databases	Data is collected, entered, and updated monthly. ☐	Data is collected, entered, and updated weekly or daily. ☐	Databases are updated as data is collected in real time ☐
Harmonization			
Refers to having common demographics (e.g. who, what, where, when variables) across datasets that	Demographics are unique to the work function and cannot be linked to data from ☐	Demographics allow for data to be connected at the division/department/crew ☐	Demographics allow for data to be connected at individual or task level ☐
Foundational Infrastructure	Refers to organizational factors that affect the ability to run analytics.		
Personnel			
Refers to the availability of key personnel with the necessary expertise to carry out technical processes of working with big data	The organization does not have employees with skills to manage large datasets and run analyses beyond Excel ® ☐	The organization has personnel skilled in managing large data and running analyses but the personnel is limited to functions like IT or ☐	Leadership engages in the formulation of business analytics questions, and has resourced an expert to work across ☐
Centralized Database			
Refers to the degree to which data variables are stored or can be readily combined into a central database	Different functions have their own various databases, but nothing is centralized ☐	Each function has their own centralized database ☐	Data across the organization is centralized, or our company uses technology that can perform a single query to all databases ☐
Measurement Culture	Measurement culture refers to the extent to which data collectors are willing to provide valid accounts of what is happening in the		
Employee Participation			
The extent to which employees participate in the process and reporting of safety matters	Employees never participate in reporting ☐	Employees sometimes report to supervisors or safety coordinators but do no direct ☐	Employees fill out reports on safety issues and participate actively in the safety process ☐
Management Use			
Support and encourage employees to participate, includes transparency about the purpose of reporting	Managers do not talk about safety or encourage participants to get involved ☐	Managers talk about safety reports or encourage participation, but don't do both ☐	Managers talk about safety reports and encourage participation ☐


ool
Diagnostic Graphics
Rules and Operations
Ready
Needs Improvement
Do Not Have
Adequate
Velocity
Harmonization
Foundational Infrastructure
Ready
Needs Improvement
Do Not Have
Personnel
Centralized Database
Measurement Culture
Ready
Needs Improvement
Do Not Have
Employee Participation
Management Use

Data Analytics Readiness Tool

Variable List	Org Metrics	Validity		
		Refers to the extent to which measures accurately represents the "real world" phenomenon targeted.		
		Do Not Have	**Needs Improvement**	**Ready**
		errors are found (e.g. missing values or impossible values, like negative hours worked)	measurement error is present, (e.g. there are outliers of some occurrences that do not seem to be plausible)	the organization has validated that the data collected represents the phenomenon it purports to measure with no errors.
Production			**Production**	
Volume Trends		☐	☐	☐
Scheduled Events		☐	☐	☐
Staffing Loads	Overtime	☐	☐	☐
Calendar Events		☐	☐	☐
Quality		☐	☐	☐
Cost/Budget		☐	☐	☐
Maintenance			**Maintenance**	
Failures (Equipment)		☐	☐	☐
Action Item Backlog		☐	☐	☐
Preventative Maintenance		☐	☐	☐
Procedures			**Procedures**	
Change Management		☐	☐	☐
Human Resources			**Human Resources**	
Turnover		☐	☐	☐
Absenteeism		☐	☐	☐
Employee Characteristics		☐	☐	☐
Culture		☐	☐	☐
Safety Metrics			**Safety Metrics**	
Hazard IDs		☐	☐	☐
Audits and Inspections		☐	☐	☐
First Aid/Minor Injury Reporting		☐	☐	☐
Near Miss/ Close Call Reporting		☐	☐	☐
Environmental		☐	☐	☐
Behavioral Observations		☐	☐	☐
Fatigue		☐	☐	☐
Safety Participation		☐	☐	☐
Safety Culture		☐	☐	☐

Data Quality

High-quality data are foundational for analyzing and using big data to realize value .

Reliability

Refers to consistency of measurement across time and units

Do Not Have	Needs Improvement	Ready
data representing the same phenomenon are recorded differently and have different definitions	data that represent the same phenomenon are recorded differently, despite having the same definition	data representing the same phenomenon have the same definition and are recorded the same
Production		
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
Maintenance		
☐	☐	☐
☐	☐	☐
☐	☐	☐
Procedures		
☐	☐	☐
Human Resources		
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
Safety Metrics		
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐

Variability

Refers to the ability of a measure to detect differences across time and units

Do Not Have	Needs Improvement	Ready
Measures tend to get the same score all the time	The measurement is sensitive enough to measure when something out of the ordinary happens	The measure is sensitive enough to detect the differences between everything commonly occurring
Production		
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
Maintenance		
☐	☐	☐
☐	☐	☐
☐	☐	☐
Procedures		
☐	☐	☐
Human Resources		
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
Safety Metrics		
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐
☐	☐	☐

Variable	Validity	Reliability	Variability	Adequate Coverage	Velocity
Production					
Volume Trends	Needs Improvement	Do Not Have	Needs Improvement	Needs Improvement	Needs Improvement
Scheduled Events	Needs Improvement	Do Not Have	Needs Improvement	Needs Improvement	Needs Improvement
Staffing Loads	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Calendar Events	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Quality	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Cost/Budget	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Maintenance					
Failures (Equipment)	Ready	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Action Item Backlog	Ready	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Preventative Maintenance	Ready	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Procedures					
Change Management	Ready	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Human Resources					
Turnover	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Absenteeism	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Employee Characteristics	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Culture	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement	Needs Improvement
Safety Metrics					
Hazard IDs	Needs Improvement	Ready	Do Not Have	Needs Improvement	Needs Improvement
Audits and Inspections	Needs Improvement	Ready	Do Not Have	Needs Improvement	Needs Improvement
First Aid/Minor Injury Reporting	Needs Improvement	Ready	Do Not Have	Needs Improvement	Needs Improvement
Near Miss/ Close Call Reporting	Needs Improvement	Ready	Needs Improvement	Needs Improvement	Needs Improvement
Environmental	Needs Improvement	Ready	Needs Improvement	Needs Improvement	Needs Improvement
Behavioral Observations	Needs Improvement	Ready	Needs Improvement	Needs Improvement	Needs Improvement
Fatigue	Needs Improvement	Ready	Needs Improvement	Needs Improvement	Needs Improvement
Safety Participation	Needs Improvement	Ready	Needs Improvement	Needs Improvement	Needs Improvement
Safety Culture	Needs Improvement	Ready	Needs Improvement	Needs Improvement	Needs Improvement

Harmonization	Personnel	Centralized Database	Employee Participation	Management Use
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready
Needs Improvement	Needs Improvement	Needs Improvement	Ready	Ready

	Necessary Capabilities for the Levels of Analytics				
	Descriptive	Diagnostic	Predictive	Prescriptive	Your Scores
Validity/Accuracy	1	1	2	2	1
Consistency	1	1	2	2	1
Variance	1	1	2	2	1
Adequate Coverage	1	1	1	2	1
Velocity		1	1	2	1
Harmonization		1	2	2	1
Personnel		1	2	2	1
Centralized Database			1	2	1
Employee Participation					2
Management Use					2
				Total Readiness Score:	58%

How ready are you to engage in these levels of analytics?			
Descriptive	Diagnostic	Predictive	Prescriptive
87%	98%	31%	25%

APPENDIX E

Company B Sample Data Excel © Sheet Cataloguing Available Data

Categories	Variables	Description	Measure/Metric	Location	Point Person	Elevate Gatekeeper	App State Keymaster	Data Quality Estimation	Notes
	Personnel								
Work Shifts	Number of hours	hours for reports, by location, coded depts	Monthly by location and by departments	Payroll or Corporate/ Kronos	Different for each facilty	[redacted]	[redacted]		Should be easy (KOD)
	Days on	hours for reports, by location, coded depts	Monthly by location and by departments	Payroll or Corporate/ Kronos	[redacted]				
	Weekends work	hours for reports, by location, coded depts	Monthly by location and by departments	Payroll or Corporate/ Kronos					
	Overnight shifts	hours for reports, by location, coded depts	Monthly by location and by departments	Payroll or Corporate/ Kronos					
	Team or individual	Different per plant							
	Vacation	hours for reports, by location, coded depts	Monthly by location and by departments	HR					
	FMLA	hours for reports, by location, coded depts	Monthly by location and by departments	HR					
	Absences (sick days)	hours for reports, by location, coded depts	Monthly by location and by departments	HR					
	Sick on the job	**Data not collected**							
	Days worked in a row	hours for reports, by location, coded depts	Monthly by location and by departments	Payroll or Corporate/ Kronos					Just switched over to new system
Job Type	Hourly/salary								
	Production worker (incentive)								
Worker Characteristics	Tenure			HR- Ulti-Pro					
	Age			HR- Ulti-Pro					
	M/F			HR- Ulti-Pro					
	Height/weight	NA							
	Distance from work	Would need to ask (have address info)							
	Demographics			HR- Ulti-Pro					
	Org commitment								
	Number of committees on	Currently NA							
Supervision	Change in supervision			HR- Ulti-Pro					
	Tenure of supervisor			HR- Ulti-Pro					
	Training supervisors			HR- Ulti-Pro					
	Number of direct reports (manager to employee ratio)			HR- Ulti-Pro					
Disciplinary	Disciplinary actions			HR- Ulti-Pro					
Employee Training Activities	Safety		sign in sheets		Safety coordinators (individual plants)	[redacted]			
	Equipment								

APPENDIX F

Measurement Culture Survey

Table 19

Summary of Measurement Culture Survey Questions and Behaviorally Anchored Rating Scale

Questions with 1-5 BARS Scale

Q1: My supervisor responds quickly to solve problems when safety issues are reported.
[1] I have not seen or heard any action being taken.
[2] I remind my supervisor multiple times before seeing any changes.
[3] I am told that action is being taken, and sometimes see results.
[4] I see that action is being taken.
[5] I typically see action within a week of reporting safety concerns.

Q2: Supervisors have us report safety-related issues to keep people safe, instead of using them solely as a performance measure.

[1] Supervisors don’t have us report safety related issues.
[2] Supervisors only ask us to report safety issues because they are told to by their boss.
[3] I don’t know what happens after I report safety issues.
[4] Supervisors explain how the reports of safety issues will be used to prevent injuries.
[5] We know how reporting safety issues helps prevent injuries and fatalities.

Q3: I report all minor injuries

[1] I never report minor injuries.
[2] I avoid reporting minor injuries.
[3] I try to report all of the minor injuries I see.
[4] When I see a minor injury, I always tell my supervisor.
[5] Every time I see a minor injury, I am sure to report it in the correct way.

Q4: I report all near misses

[1] I am always too busy to report near misses to peers, safety coordinators, or supervisors.
[2] I am too busy to stop work to report near misses unless I think they could result in a serious injury.
[3] I try to report all of the minor injuries I see.
[4] Even though I am busy, I tell my supervisor about near misses so they will be recorded.
[5] Every time I see a near miss, I am sure to report it in the correct way.

Q5: I find the forms used to report safety information easy to use.

[1] I do not complete forms to report safety issues.
[2] I don’t fill out forms, but I tell my supervisor so that they will document what I share.
[3] Sometimes I use forms to report safety concerns.
[4] Whenever I need to report concerns, I try to fill out a safety form.
[5] I am able to find, navigate, and submit all safety forms with ease.

Questions with 1-5 BARS Scale
Q6: My supervisor encourages employees to participate in decisions which affect safety (operating procedures, PPE…) [1] I'm never asked decide anything around safety. [2] I never get asked but my supervisor tells us about safety decisions. [3] I get asked for my opinion but I rarely give it. [4] My supervisors encourage my opinion but am not involved in any final decisions. [5] Me and my work team make decisions about our own safety. **Q7: I am involved in safety audits, inspections, and behavior observations on a regular basis.** [1] I have never been asked to do this. [2] I have the opportunity to do these but chose not to. [3] I do these only because they are mandatory each month. [4] I volunteer to do one of these at least once a month. [5] I voluntarily participate in these as much as able. **Q8: My supervisor regularly asks employees about safety concerns and listens to our ideas** [1] My supervisor does not listen when we bring up our concerns. [2] My supervisor seems to listen to our safety concerns but nothing happens. [3] I am comfortable approaching my supervisors about any concerns I have about safety. [4] My supervisor wants to hear about safety concerns at every meeting and writes them down. [5] I can freely talk to my supervisor anytime about safety concerns and things get done. **Q9: My supervisor talks about lessons from incidents and other things we've reported (minor injuries, near misses).** [1] My supervisor never talks about safety things that happened around the plant. [2] My supervisor just reads findings to my team when required. [3] My supervisor will talk about what we've learned from incidents when they know results. [4] My supervisor asks us what we want to learn about based on what has been reported. [5] My supervisor asks us to look for things we've learned to talk about as a crew. **Q10: Improvements made because of our safety reporting.** [1] I'm not aware of any safety improvements recently [2] Supervisors will talk about improvements but they are not that big. [3] We get safety improvements but I don't see how they are related to what I report. [4] Improvements are made, because of how our team reports safety issues. [5] We get praised when the things we've reported improves our safety. **Q11: I help investigate safety incidents and near misses.** [1] I avoid being involved in investigations because I am afraid I may get in trouble or look stupid.

Questions with 1-5 BARS Scale
[2] I don’t consider this part of my job. [3] I cooperate with investigations when required. [4] I voluntarily participate in investigations when I think I can help. [5] I encourage my peers to also get involved in investigations of safety incidents. **Q12: All incidents that have the potential for serious injury (P-SIFs) are thoroughly investigated with accurate information.** [1] I really doubt it, most of us hide incidents and close calls. [2] My supervisor collects information, but doesn't ask the right questions. [3] I have seen supervisors collect information, but only the bare-minimum. [4] My supervisors collects thorough information around incidents. [5] My supervisors asks everyone involved to correct any errors in the investigation. **Q13: There is so much "pencil whipping" (completing the form without doing inspection or observation) that data quality cannot be trusted.** [1] Most of us fill out paperwork without doing the requested inspection or observation. [2] Sometimes I make up something afterward instead of filling out paperwork in the moment. [3] I wait until the end of the day or a break to fill out safety paperwork. [4] I complete paperwork immediately but very quickly and do not double-check the forms. [5] I am careful to accurately add all relevant information as soon as possible. **Q14: Safety audits, inspections, and observations are routinely performed in my work area.** [1] These are never done in my work area. [2] These are only done when they are required or if there has been recent incident. [3] I notice these being done but they seem to only happen at the end of the month. [4] I consistently see these being done and I routinely participate in them. [5] We talk about these in our team meetings and encourage each other to get involved.

APPENDIX G

IRB Approval

Maira Compagnone <compagnoneme@appstate.edu>

IRB Notice - 20-0059

IRB <irb@appstate.edu> Thu, Sep 19, 2019 at 4:07 PM
To: compagnoneme@appstate.edu, ludwigtd@appstate.edu

To: Maira Compagnone
Psychology , College of Business
CAMPUS EMAIL

From: IRB Administration
Date: 9/19/2019
RE: Determination that Research or Research-Like Activity does not require IRB Approval

STUDY #: 20-0059
STUDY TITLE: A measurement of organizational readiness for analytics within occupational health and safety

The IRB determined that the activity described in the study materials does not constitute human subject research as defined by University policy and the federal regulations [45 CFR 46.102 (d or f)] and does not require IRB approval.

This determination may no longer apply if the activity changes. IRB approval must be sought and obtained for any research with human participants.

If you have any questions about this determination, please contact Robin Tyndall at 262-2692; or irb@appstate.edu.
Thank you.

Appendix H

DART PDF Manual

Data Analytics Readiness Tool

Organization Name/Level

Date

Purpose of Readiness Assessment 3

Description of Levels of Analytics 4

Descriptive Analytics 4

Diagnostic Analytics 4

Predictive Analytics 4

Prescriptive Analytics 4

DART Success Factors: 5

What are we assessing, and why? 5

Rules and Operations (O): 5

Foundational Infrastructure (I): 5

Measurement Culture (C): 5

Data Quality (Q): 5

Rating Data Quality Using the Variable List and Measurement Framework 7

What is the Variable List? 7

Variable List Workbook 8

Reference the Measurement Framework 9

Data Analytics Readiness Tool (DART) Directions 10

Sheet 1: Organizational Level Ratings 10

Sheet 2: Data Quality Variable Ratings 11

Sheet 3: Data Matrix Summary 12

Sheet 4: Aggregate Readiness Scores 13

Purpose of Readiness Assessment

The safety industry can benefit greatly from the implementation of data analytics, but often don't have all of the necessary components in place. A readiness assessment is needed for organizations to be able to reliably predict the expected process maturity with a guide for areas that need to be further developed. The purpose of this functional readiness assessment tool for the safety industry is to diagnose current analytic capabilities and provide guidance on how to improve across four success factors: a) rules and operations, b) foundational infrastructure c) measurement culture and d) data quality.

Description of Levels of Analytics

There are four levels of analytics that can be applied using data: a) descriptive analytics, b) diagnostic analytics, c) predictive analytics, and d) prescriptive analytics.

Descriptive Analytics

Descriptive analytics answer questions about what has happened in the past. Within OHS, safety data and information is analyzed for characteristics and relationships (e.g. similarity) through descriptive statistics and data visualizations (Huang, Wu, Wang, & Ouyang, 2018). Examples of this type of analytics include sums, means, and averages. These are used to clarify and define the current safety state of an organization through reports and dashboards.

Diagnostic Analytics

Diagnostic analytics provide clues as to the reason for such past occurrences (e.g., Why did it occur?). These types of analytics (e.g. correlational analysis) use relationships between variables to provide context. Diagnostic analytics use historical and past safety performance to identify reasons for the success or failure of initiatives, or the reason for the occurrence of specific safety outcomes by investigating causal relationships, outliers, and sequences (Huang et al., 2018).

Predictive Analytics

Predictive analytics (e.g. regression analysis) attempts to forecast future outcomes (e.g., What will happen and why?). In addition to historical or past data, predictive analysis incorporates current information in an attempt to predict the likelihood of a situation occurring (Huang et al., 2018). Predictive analytics in OHS can span temporally from short-term predictions to long-term.

Prescriptive Analytics

Finally, prescriptive analytics take predictive analytics and utilizes data streams that are updated in real-time in order to provide the most accurate guidance for decision-making (e.g., What should I do and why should I do it?; Mousanif, Saba, Douiji, & Sayad, 2014). Safety data, mathematical formulae, safety rules, and machine learning are synchronously run to suggest the most advantageous decision options, based on the identification of future opportunities or risks and the implications of each option (Huang et al., 2018).

DART Success Factors:

What are we assessing, and why?

In order to run successful analytics, there are certain components that experts have identified as necessary. These categories are called "success factors" because high ratings of readiness in these categories will allow you to run successful analytics.

Rules and Operations (O):

In assessing the rules and operations of how we collect, store, and manage data, we want to be sure that we are capturing as many variables of interest as possible: variables in the Variable List and outcome variables. It is also important to consider how often the data is collected and entered into databases. The operational definitions (e.g. demographic naming conventions that refer to who, what, where, or when) of these variables are assessed for consistency.

Foundational Infrastructure (I):

Factors involving infrastructure include if and how the data are shared across organizational divisions and functions (e.g., safety office, production, human resources), and assessing how the data are stored. Additionally, the personnel support needed to manage and clean the data must be assessed, along with strategic leadership and an ability to make meaningful decisions based on analytical findings.

Measurement Culture (C):

Employee perceptions of and willingness to engage in safety measurement might affect the validity and availability of the data. A company can have the best measurement infrastructure possible, but this measurement system will be ineffective for analysis and improvement if employees are not willing to participate in that system by completing forms, conducting observations, or reporting close-call incidents accurately and timely. Accordingly, employee participation in the collection and reporting of information, the quality of employee-shared data, and the managers' receptivity to employee feedback and communication of how data relates to informed decision-making will be assessed under this factor.

Data Quality (Q):

Conducting a data audit and data cleaning, while often the least glamorous analytics phase, is often the most critical. This is especially the case with pre-existing databases. Because pre-existing databases are often created for purposes other than data analysis, the quality of the data will be driven by what was important in the original use of this data and hence need not satisfy the quality requirements for the proposed analyses. This includes assessing the validity, reliability, and variance of the data. Steps in this objective will need to occur any time new information is collected and added to the centralized database.

Rating Data Quality Using the Variable List and Measurement Framework

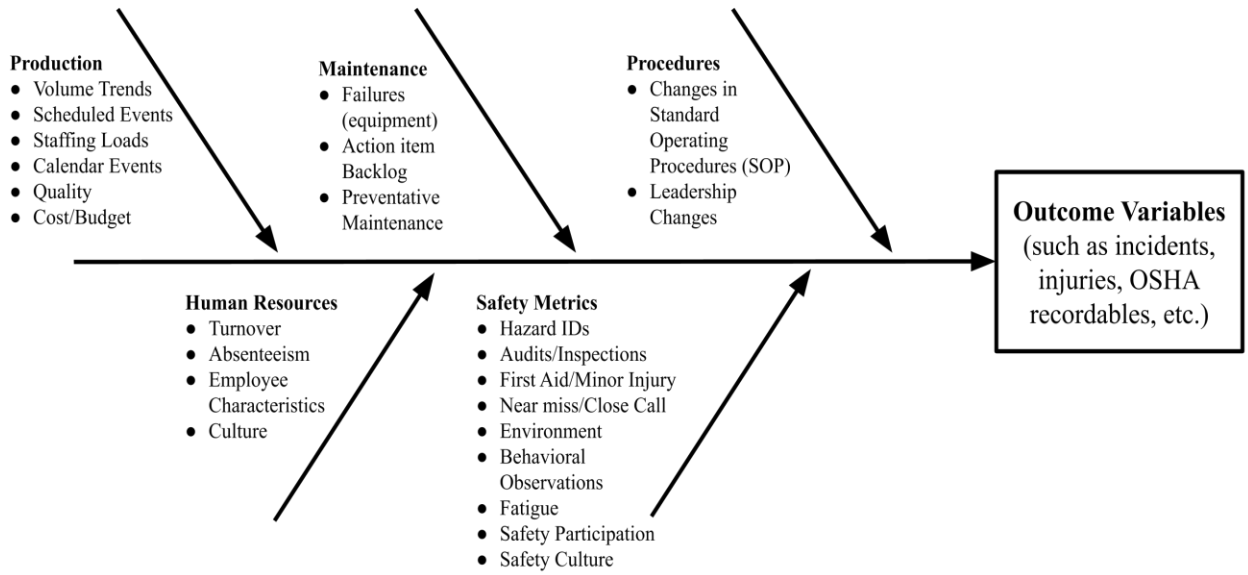
What is the Variable List?

The Variable List has been assembled based on the metrics that organizations collect that research and safety professionals have identified as important to predict safety outcomes. Each metric falls under one of 5 areas of reporting. Often organizations have their own names for things, so what your company collects for “staffing loads” may look like “number of workers on a scheduled day”, or “overtime hours by crew, weekly”. Use the workbook on the next page to write down what metrics you currently collect for that variable in the corresponding blocks under each category.

Reference the Measurement Framework

The Measurement Framework shows the relationships that each variable in the Variable List has to the outcome variables. This provides a holistic picture of why it is important to have adequate coverage of these variables in order to assess what may be causing injuries and fatalities. Consider what variables your organization may not be collecting information on.

Next, we will turn to a comprehensive list of all of these variables that may be impacting safety outcomes in your organization, and you will have the opportunity to input the metric names for the variables that you collect data on.



Variable List Workbook

Variable	Definition	Your list of applicable metrics
Production		
Volume Trends	The quantity of product output at a given time.	
Scheduled Events	These metrics relate to changes to the production that managers may make, such as switching a machine from one product to another.	
Staffing Loads	These metrics relate to accounting for the amount of people working at a given time, and how much those employees are working.	
Calendar Events	These are events that may center around a date on a calendar, like a holiday or vacation.	
Quality	A measure of excellence or a state of being free from defects, deficiencies and significant variations. These metrics may also measure the opposite, such as errors or reworks.	
Cost/Budget	These metrics would measure estimated costs, revenues, and resources over a specified period.	
Maintenance		
Failures (Equipment)	These metrics track what machines have failed, how often, or why.	
Action Item Backlog	These metrics measure unfinished tasks that need to be completed.	
Preventative Maintenance	Data centered around equipment and facilities by tracking systematic inspection, detection, and correction of incipient failures either before they occur or before they develop	

	into major defects.	
Procedures		
Change Management	These metrics could track what occurs after a change in procedure, product, or could track leadership changes over time. These organizational metrics may cross-functions and help to provide a strategic lens.	
Human Resources		
Turnover	A measurement of the number of employees who leave an organization during a specified time period.	
Absenteeism	The measure of the number of employees that are absent from their scheduled shift.	
Employee Characteristics	Demographics of the workforce that is being measured (e.g., tenure, sex, age, education).	
Culture	The measure of the cultural norms of the workforce, including what they prioritize, talk about, and behave.	
Safety Metrics		
Hazard ID	The metrics surrounding the reports made by employees of hazardous environments that could lead to safety incidents.	
Audits/Inspections	These metrics track the frequency of audits and inspections of workplaces and workgroups to measure if the work is being done safely and if the environment is safe.	
First Aid/Minor Injury Reporting	These metrics track minor injuries, such as small cuts, trips, or falls that may need first aid.	

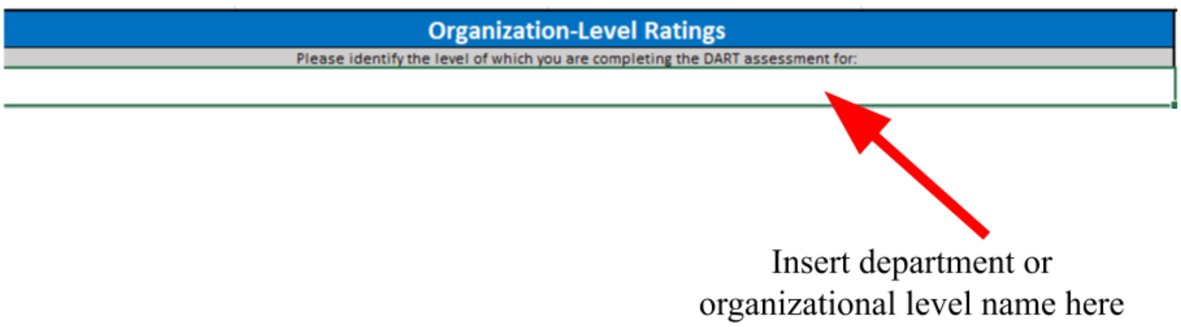
Near Miss/Close Call Reporting	This metric measures the number of instances where employees report that a safety incident did not occur, but almost did.	
Environmental	Environmental metrics track adverse conditions such as weather, heat, wind, storms, etc.	
Behavioral Observations	Metrics collected on behavioral observations may include video or checklists, either paper and pencil forms or electronic.	
Fatigue	Measures tracking the point at which employees are likely to have a safety incident due to strain and fatigue caused by the nature of their work/workload.	
Safety Participation	Measures of safety participation track whether front line employees are engaged with the safety process, including shift discussions, reporting, or investigating incidents.	
Safety Culture	Safety culture measures the extent to which employees and managers think about, talk about, and are committed to having a safe workplace.	

Data Analytics Readiness Tool (DART) Directions

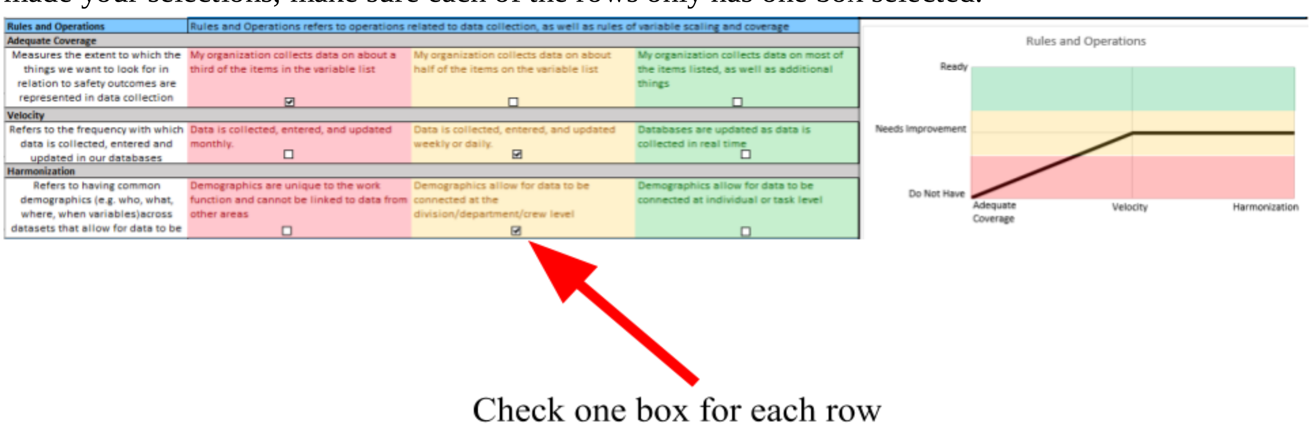
Open the DART Excel file and follow the directions listed below for each excel worksheet in order.

Sheet 1: Organizational Level Ratings

The DART is designed to assess readiness for analytics at an organizational level or for specific departments if the departments vary in their data collection processes. Different data collection processes will lead to different overall analytics readiness ratings, so before you begin filling out one DART for multiple departments, make sure their data collection processes are consistent. To track which level or department you are assessing, type it in under “Organization-Level Ratings.”



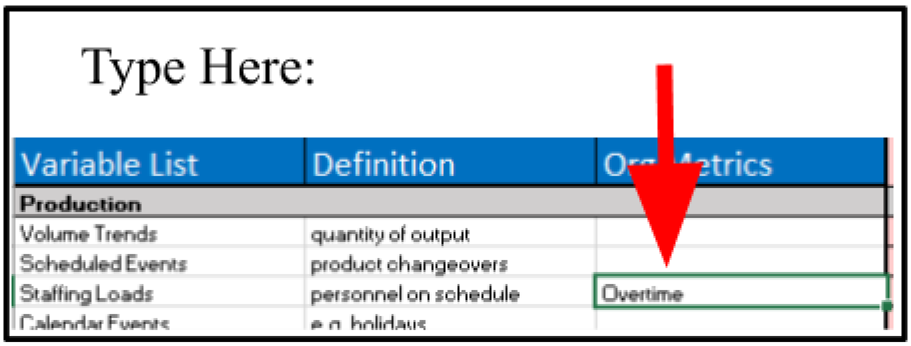
Check one box for the anchor that most accurately represents the level of data collection for each facet of the organizational success factors at the department or organizational level you have specified above. Depending on which anchors you select for each organizational success factor, it will populate a chart on the right to give you a visualization of the areas in each success factor that need to be improved. If the charts do not populate a line after you’ve made your selections, make sure each of the rows only has one box selected.

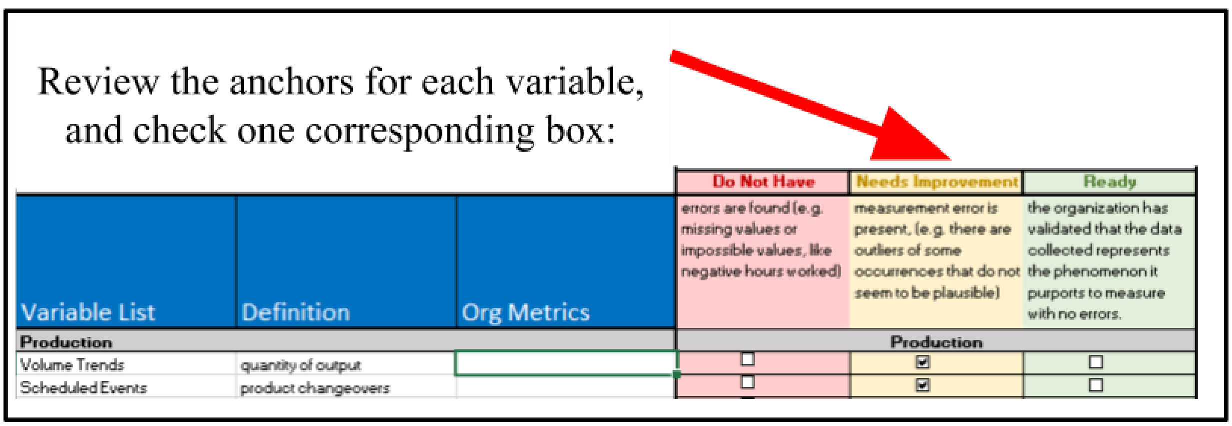


Sheet 2: Data Quality Variable Ratings

While the Org Level Ratings assess success factors that are likely common across the level, data quality will vary greatly for each metric, depending on what the variable is trying to assess, who is collecting or reporting the information, and how sensitive the measurement is to changes in context or environment. For that reason, each variable must be rated for validity, reliability, and variability separately.

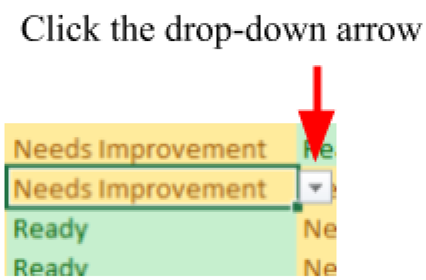
On the Excel sheet labeled "Data Quality Variable Ratings," input the metrics names from your worksheet for each variable, and carefully consider the variable against the anchors for each factor. Be sure to check only one box per line.





Sheet 3: Data Matrix Summary

We recognize that the "Org Level Ratings" apply broadly, but that one or more of the variables may fall under different circumstances. For example, you may have employee participation in reporting safety metrics, but employees do not enter human resources variables. In the Data Matrix Summary, you have the opportunity to review all of the ratings collected so far, and change any of the Org Level Ratings for individual variables that may be different.

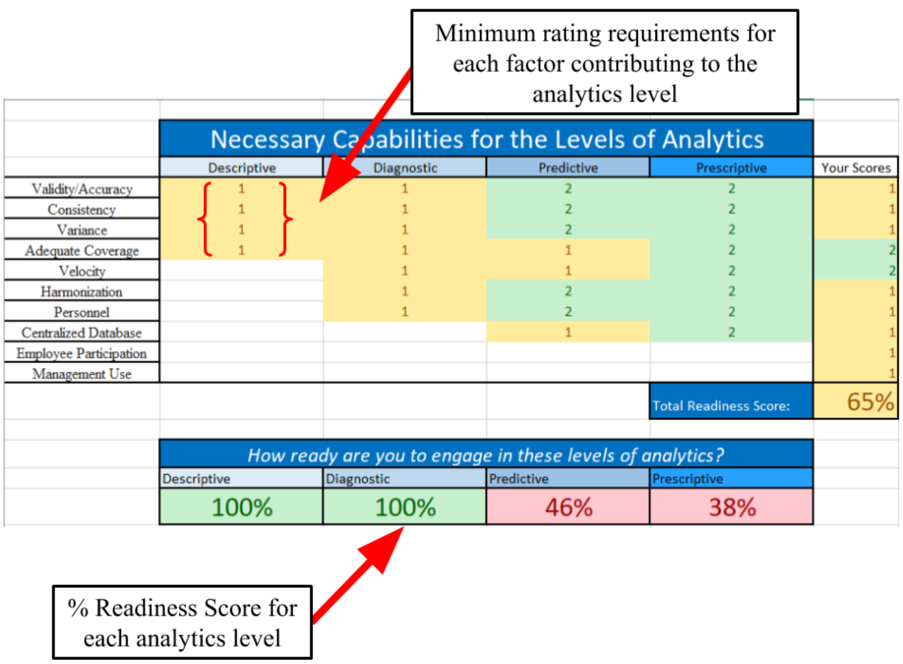


Sheet 4: Aggregate Readiness Scores

When you turn to the Aggregate Readiness Scores, your ratings will calculate into an overall readiness score, along with some information about the levels of analytics and your readiness. The square shows you information on what is necessary to run each level on analytics. The "1" denotes that at least a rating of "Needs Improvement" is necessary. "2" means that a score of "Ready" is required for that success factor. The success factors are weighted for their relative importance in assessing readiness. For example, having high quality data will influence your analytics more than having an employee participate in reporting. Your weighted scores will reflect your overall % of readiness.

For each level of analytics, your weighted scores will be compared to the minimum cut-offs, and will present you with a % readiness for each level.

The measurement culture success factor is not necessary to run analytics, but will provide valuable diagnostic information to your organization. If you score low in measurement culture, it will affect each of the other success factors. In this way, it provides valuable information for the next steps in developing your organization's capabilities.



Necessary Capabilities for the Levels of Analytics

	Descriptive	Diagnostic	Predictive	Prescriptive	Your Scores
Validity/Accuracy	1	1	2	2	1
Consistency	1	1	2	2	1
Variance	1	1	2	2	1
Adequate Coverage	1	1	1	2	2
Velocity		1	1	2	2
Harmonization		1	2	2	1
Personnel		1	2	2	1
Centralized Database			1	2	1
Employee Participation					1
Management Use					1
				Total Readiness Score:	65%

How ready are you to engage in these levels of analytics?

Descriptive	Diagnostic	Predictive	Prescriptive
100%	100%	46%	38%

APPENDIX I

Measurement Culture Survey Summary Statistics

Table 20

Descriptive Statistics for Company A

Questions	N	Mean	Median	Std. Deviation	Skewness	Kurtosis	Std. Error
Q1	348	4.50	5.00	0.75	-1.60	2.49	0.04
Q2	348	4.49	5.00	0.73	-1.59	3.01	0.04
Q3	348	4.27	5.00	1.12	-1.41	0.90	0.06
Q4	348	4.29	5.00	0.93	-1.35	1.58	0.05
Q5	348	3.30	3.00	1.11	-0.23	-0.52	0.06
Q6	348	4.34	5.00	0.82	-1.24	1.42	0.04
Q7	348	3.64	4.00	1.16	-0.59	-0.44	0.06
Q8	348	4.44	5.00	0.77	-1.46	2.34	0.04
Q9	348	4.26	4.00	0.87	-1.25	1.59	0.05
Q10	348	3.78	4.00	0.90	-0.28	-0.48	0.05
Q11	348	2.94	3.00	1.26	0.01	-0.96	0.07
Q12	348	4.40	5.00	0.80	-1.26	1.26	0.04
Q13[a]	348	2.49	2.00	1.10	0.40	-0.55	0.06
Q14	348	4.14	4.00	0.86	-1.07	1.32	0.05

Note: Table displaying statistics for the measurement culture survey across Company A's Fibers Division.
[a]Denotes reverse coded responding.

Table 21

Descriptive Statistics for Company B

Questions	N	Mean	Median	Std. Deviation	Skewness	Kurtosis	Std. Error
Q1	145	3.90	4.00	0.79	-1.26	2.93	0.05
Q2	148	4.33	4.00	0.75	-1.18	1.36	0.06
Q3	152	3.75	4.00	1.26	-0.90	0.35	0.09
Q4	149	3.88	4.00	1.07	-1.01	0.60	0.07
Q5	148	3.51	4.00	1.33	-0.41	-1.17	0.08
Q6	151	3.95	4.00	1.05	-1.12	0.71	0.06
Q7	137	3.27	3.00	1.49	-0.27	-1.25	0.08
Q8	149	3.82	4.00	0.98	-0.37	-0.13	0.06
Q9	143	3.59	3.00	1.05	-0.15	-0.45	0.06
Q10	144	3.85	4.00	1.00	-1.64	2.62	0.06
Q11	142	3.26	3.00	0.83	0.29	-0.03	0.09
Q12	143	4.16	4.00	0.76	-1.69	5.30	0.05
Q13	132	4.42	5.00	1.07	-1.98	3.27	0.08
Q14	147	4.02	4.00	1.13	-0.87	0.24	0.06

Note: Table displaying statistics for the measurement culture survey across Company B

APPENDIX J

DART Reports for Recommendations (Company A & B)

DART REPORT

Data Analytics Readiness Tool

Appalachian State University and [redacted] Company

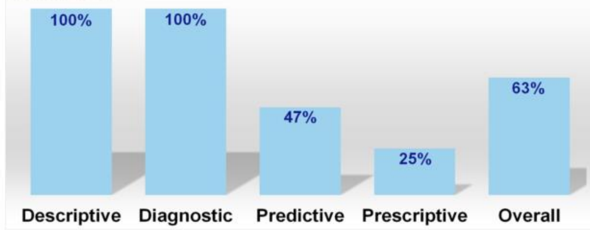
APRIL 2020

PREPARED BY MAIRA COMPAGNONE & AVA YOUNG

Report Content Overview

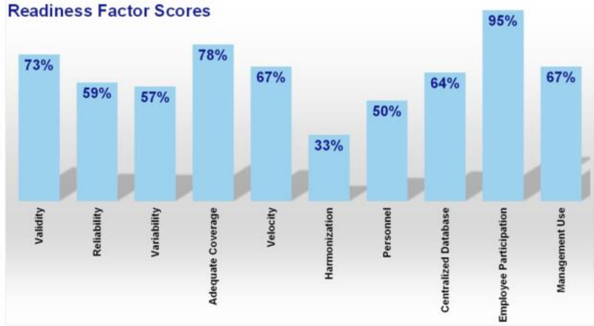
DART Results | Readiness Factor Scores | Measurement Culture Survey | Action Steps

DART Results



DART self-assessments were averaged across all respondents. The self-assessment recommends focusing efforts on improving the measurement system in order to maximize **PREDICTIVE ANALYTICS** capabilities.

Readiness Factor Scores



Readiness factors over 75% are considered optimal, and readiness factors below 50% are considered sub-optimal. In order to maximize predictive analytics [redacted]ical Company should (a) ensure the availability of common variables that can be used to connect data (**HARMONIZATION**), (b) dedicate personnel that can clean, aggregate, and analyze safety data (**PERSONNEL**), and (c) improve the sensitivity of safety measurement systems to allow them to capture more variance (**VARIABILITY**).

Average Quality per Variable

Variables That Need Improvement

Validity: measures accurately represents the "real world" item

- Fatigue

Reliability: consistency of measurement across time and units

- Fatigue
- Preventative Maintenance
- Change Management

Variability: measures detect differences across time and units

- Fatigue
- Preventative Maintenance
- Change Management
- Behavioral Observations
- Safety Participation
- Safety Culture
- Culture

DART REPORT

Appalachian State University ar [redacted] Company

Measurement Culture Survey

Measurement Culture will impact Data Quality. [redacted] mpany should focus on improving measurement culture by (a) reevaluating checklists and forms to ensure they are simple and easy to use, which will also decrease pencil-whipping, and (b) involving employees in all areas of reporting.

The culture survey supported the DART Measurement Culture Self-Assessment.

Focus on improving Management Action by:

- Increasing transparency of the purpose of data collection
- Showing employees how data is used to make better safety decisions

Action Step Recommendations

OPPORTUNITIES FOR MAXIMUM IMPROVEMENT

Improvements to Readiness Factors that are under 50% optimized will have the biggest impact on analytics readiness.

- **Harmonization**: In order to maximize safety analytics [redacted] Company must ensure that common variables exist across databases. This will allow disparate databases to be linked together to find predictive relationships.
- **Personnel Infrastructure**: In order to maximize safety analytics, [redacted] Chemical Company must dedicate manpower and additional resources to cleaning, aggregating, and running analyses on safety variables.

AREAS OF STRENGTH TO MAINTAIN

Improvements to Readiness Factors that are over 75% optimized will be the easiest (e.g. lower cost or effort) areas to improve analytics readiness.

- **Adequate Coverage**: [redacted] Company has metrics on many of the essential predictors of safety outcomes. Continue to improve the data quality of these measures to enhance the value of predictive analytics.
- **Employee Participation**: Employee participation in reporting will affect data quality. Continue to support and encourage employees to accurately report safety incidents in order to improve data quality of those measures.

Appalachian State University Research Team
Dr. Yalcin Acikgoz, Dr. Shawn Bergman, & Dr. Tim Ludwig
Rachel Bellflowers, Sam Biggs, Julie Brooks, Maira Compagnone, Cori Ferguson, Nicholas Garnowski, Philip Hinson, Matthew Laske, Connor Linden, Royale Nicholson, Tara O'Neil, & Ava Young

DART REPORT

Data Analytics Readiness Tool

Appalachian State University [redacted]

APRIL 2020

PREPARED BY MAIRA COMPAGNONE & AVA YOUNG

Report Content Overview

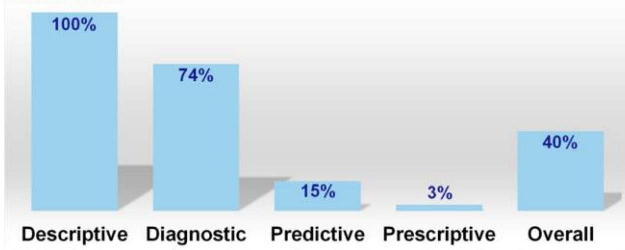
DART Results

Readiness Factor Scores

Measurement Culture Survey

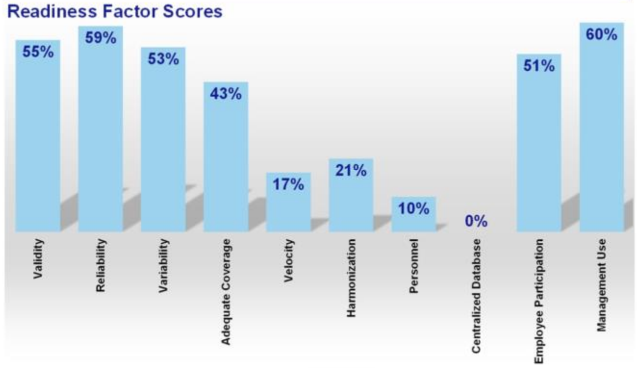
Action Steps

DART Results



DART self-assessments were averaged across all respondents. The self-assessment recommends focusing efforts on improving the measurement system in order to maximize ***DIAGNOSTIC ANALYTICS*** capabilities.

Readiness Factor Scores



In order to maximize predictive analytics, [redacted]hould (a) create a central database where data variables can be stored or readily combined (**CENTRALIZED DATABASE**), (b) dedicate personal that can aggregate, and analyze safety data (**PERSONNEL**), and (c) increase the frequency by which data is collected, entered, and updated (**VELOCITY**).

Average Quality per Variable

Variables That Need Quality Improvements

<u>Validity</u>: measures accurately represents the "real world" item

<u>Reliability</u>: consistency of measurement across time and units

<u>Variability</u>: measures detect differences across time and units

- Safety Participation
- Fatigue
- Behavioral Observations
- Environmental
- Culture
- Change Management
- Failures (Equipment)
- Scheduled Events

DART REPORT

Appalachian State University and [redacted]

Measurement Culture Survey

Measurement Culture will impact Data Quality. [redacted] uld improve Measurement Culture by involving employees in investigations, safety audits, inspections, and behavior observations.

The front-line culture survey showed that employees want more Management Action Items (Q6-14).

Focus on improving Management Action by:

- Showing employees how data is used to make better safety decisions
- Increasing transparency of the purpose of data collection
- Ensuring ease of use of safety forms

Action Step Recommendations

OPPORTUNITIES FOR MAXIMUM IMPROVEMENT

Improvements to Readiness Factors that are under 50% optimized will have the biggest impact on analytics readiness.

- **Centralized Database**: In order to maximize safety analytics [redacted] ust create a centralized database where data variables can be stored or readily combined.
- **Personnel**: In order to maximize safety analytics, [redacted] st dedicate manpower and additional resources to cleaning, aggregating, and running analyses on safety variables.

AREAS OF STRENGTH TO MAINTAIN

Improvements to Readiness Factors that are currently more optimized will be the easiest (e.g. lower cost or effort) areas to improve analytics readiness.

- **Management Use**: [redacted] anagement shows employees how data is used to find problems and make positive changes. Continue to improve in this area of Measurement Culture to enhance the value of Diagnostic Analytics.
- **Reliability**: [redacted] maintains the consistency of measurement across time and units. Continue to improve in this area of data quality to enhance the value of Diagnostic Analytics.

Appalachian State University Research Team
Dr. Yalcin Acikgoz, Dr. Shawn Bergman, & Dr. Tim Ludwig
Rachel Bellflowers, Sam Biggs, Julie Brooks, Maira Compagnone, Cori Ferguson, Nicholas Garnowski, Philip Hinson, Matthew Laske, Connor Linden, Royale Nicholson, Tara O'Neil, & Ava Young

DART REPORT

GLOSSARY OF TERMS

Absenteeism	The measure of the number of employees that are absent from their scheduled shift.
Action Item Backlog	These metrics measure unfinished tasks that need to be completed.
Adequate Coverage	Measures the extent to which the things we want to look for in relation to safety outcomes are represented in data collection
Audits/ Inspections	These metrics track the frequency of audits and inspections of workplaces and workgroups to measure if the work is being done safely and if the environment is safe.
Behavioral Observations	Metrics collected on behavioral observations may include video or checklists, either paper and pencil forms or electronic.
Calendar Events	These are events that may center around a date on a calendar, like a holiday or vacation.
Centralized Database	Refers to the degree to which data variables are stored or can be readily combined into a central database
Change Management	These metrics could track what occurs after a change in procedure, product, or could track leadership changes over time. These organizational metrics may cross-functions and help to provide a strategic lens.
Cost/Budget	These metrics would measure estimated costs, revenues, and resources over a specified period.
Culture	The measure of the cultural norms of the workforce, including what they prioritize, talk about, and behave.
Employee Characteristics	Demographics of the workforce that is being measured (e.g., tenure, sex, age, education).
Employee Participation	The extent to which employees participate in the process and reporting of safety matters
Environmental	Environmental metrics track adverse conditions such as weather, heat, wind, storms, etc.
Failures (Equipment)	These metrics track what machines have failed, how often, or why.
Fatigue	Measures tracking the point at which employees are likely to have a safety incident due to strain and fatigue caused by the nature of their work/ workload.
First Aid/Minor Injury Reporting	These metrics track minor injuries, such as small cuts, trips, or falls that may need first aid.
Harmonization	Refers to having common demographics (e.g. who, what, where, when variables) across datasets that allow for data to be linked
Hazard ID	The metrics surrounding the reports made by employees of hazardous environments that could lead to safety incidents.
Management Action	Support and encourage employees to participate, includes transparency about the purpose of reporting
Near Miss/Close Call Reporting	This metric measures the number of instances where employees report that a safety incident did not occur, but almost did.
Personnel Infrastructure	Refers to the availability of key personnel with the necessary expertise to carry out technical processes of working with big data
Preventative Maintenance	Data centered around equipment and facilities by tracking systematic inspection, detection, and correction of incipient failures either before they occur or before they develop into major defects.
Quality	A measure of excellence or a state of being free from defects, deficiencies and significant variations. These metrics may also measure the opposite, such as errors or reworks.
Reliability	Refers to consistency of measurement across time and units
Safety Culture	Safety culture measures the extent to which employees and managers think about, talk about, and are committed to having a safe workplace.
Safety Participation	Measures of safety participation track whether front line employees are engaged with the safety process, including shift discussions, reporting, or investigating incidents.
Scheduled Events	These metrics relate to changes to the production that managers may make, such as switching a machine from one product to another.
Staffing Loads	These metrics relate to accounting for the amount of people working at a given time, and how much those employees are working.
Turnover	A measurement of the number of employees who leave an organization during a specified time period.
Validity	Refers to the extent to which measures accurately represents the "real world" phenomenon targeted
Variability	Refers to the ability of a measure to detect differences across time and units
Velocity	Refers to the frequency with which data are collected, entered and updated in our databases
Volume Trends	The quantity of product output at a given time.

Vita

Maira Elize (Ezerins) Compagnone was born in Massachusetts. She received her Bachelor of Arts in Ethical, Social, and Political Philosophy from the University of Massachusetts, Boston, in 2011. In the fall of 2018, she accepted a research assistantship in Psychology at Appalachian State University and began study toward a Master of Arts in Industrial-Organizational Psychology and Human Resources Management and a Master of Business Administration degrees. The M.A. and M.B.A were awarded in May of 2020. She plans to begin study at the University of Arkansas in the fall of 2020, where she will begin work towards her Ph.D. in Management at the Walton College of Business.