International Journal of Industrial Ergonomics 85 (2021) 103184

Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

journal homepage: www.elsevier.com/locate/ergon

Identifying step-level complexity in procedures: Integration of natural language processing into the Complexity Index for Procedures—Step level (CIPS)

Farzan Sasangohar[a,d], Nilesh Ade[b], Noor Quddus[b], S. Camille Peres[b,c,*], Pranav Kannan[b]

[a] Wm Michael Barnes '64 Department of Industrial & Systems Engineering, Texas A&M University, 3131 TAMU, 101 Bizzell Street, College Station, TX, 77843 USA
[b] Mary Kay O'Connor Process Safety Center, Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, USA
[c] ErgoCenter, Environmental and Occupational Health, Texas A&M University, College Station, TX, USA
[d] Center for Outcomes Research, Houston Methodist Hospital, Houston, TX, USA

ARTICLE INFO

Keywords:
Task complexity
Written procedure
Complexity measurement
Natural language processing
Machine learning
Artificial intelligence

ABSTRACT

Task complexity plays an important role in performance and procedure adherence. While studies have attempted to assess the contribution of different aspects of task complexity and their relationship to workers' performance and procedure adherence, only a few have focused on application-specific measurement of task complexity. Further, generalizable methods of operationalizing task complexity that are used to both write and evaluate a wide range of routine or non-routine procedures is largely absent. This paper introduces a novel framework to quantify the step-level complexity of written procedures based on attributes such as decision complexity, need for judgment, interdependency of instructions, multiplicity of instructions, and excess information. This framework was incorporated with natural language processing and artificial intelligence to create a tool for procedure writers for identifying complex elements in procedures steps. The proposed technique has been illustrated through examples as well as an application to a tool for procedure writers. This method can be used both to support writers when constructing procedures as well as to examine the complexity of existing procedures. Further, the complexity index is applicable across several high-risk industries in which written procedures are prevalent, improving the linguistic complexity of the procedures and thus reducing the likelihood of human errors with procedures associated with complexity.

1. Introduction

Safe and effective performance in high-risk industries (e.g., petrochemical, and oil and gas) depend on effective execution of tasks and often written procedures are used to facilitate this (Bullemer and Hajdukiewicz, 2004). However, there continues to be troubling issues with procedural systems as problems with these systems continue to be associated with incidents. Out of approximately 100 incidents investigated by the Chemical Safety and Hazard Investigation Board (CSB) since its inception in January 1998, 49 involved some type of procedural deficiency (Chemical Safety and Hazard Investigation Board, 2018). The BP Texas City refinery incident is an instance, where the inadequate hazard identification and improper procedure use coupled with worker fatigue led to an explosion that caused multiple fatalities and huge economic fallout (Holmstrom et al., 2006). The ExxonMobil Torrance refinery incident is another example, where inadequate hazard identification, lack of safe operating limits and procedures, and variance in procedure use were the major causes of the incident (Chemical Safety and Hazard Investigation Board, 2017). In both cases, the procedural failure either by improper use or inadequate content was a critical contributor to the incidents. This is in line with a review of thousands of maintenance reports from the NASA aviation safety reporting system which found errors in both document deficiency and user errors (Lattanzio et al., 2008).

Procedures usually convey expectations and guidance from the operational management to the worker how they expect various tasks to be completed. Ostensibly, this guidance provides a safe, logical, and efficient means of carrying out the objectives of the task (Degani and

Abbreviations: CSB, Chemical Safety and Hazard Investigation Board; NLTK, Natural Language Toolkit.

* Corresponding author. ErgoCenter, 212 Adriance Lab Road, College Station, TX 77843-1266 USA.
E-mail addresses: sasangohar@tamu.edu (F. Sasangohar), nilesh14@tamu.edu (N. Ade), nooralquddus@tamu.edu (N. Quddus), peres@tamu.edu (S.C. Peres), pranav.kannan92@gmail.com (P. Kannan).

https://doi.org/10.1016/j.ergon.2021.103184
Received 20 October 2020; Received in revised form 1 July 2021; Accepted 2 July 2021
Available online 6 August 2021
0169-8141/

Wiener, 1997). However, the complexity of most modern high-risk industrial environments makes it difficult to create procedures that always support effective, efficient, and safe behavior. Complex interactions among the workers, the task, and the process environment are expected in sociotechnical systems and may further exacerbate issues with procedures. These interactions can vary widely with the experience level of the worker, the frequency of the task, interruptions from other workers, weather conditions, and safety culture (Ahmed et al., 2020; Peres et al., 2016; Williams et al., 2017). Given the tight coupling and highly interactive nature of these systems, actions taken by workers may have consequences that propagate beyond a simple cause-effect relationship. When viewed through this complex socio-technical system lens, the challenges regarding how to design procedures to support the workers' interactions are unique and non-trivial, as currently procedures are assumed to accommodate every worker, in every situation, and that every situation is controlled and ideal. This may result in workarounds from the procedural instructions which impose additional challenges for evaluating the effectiveness of procedures. Such intentional deviations or workarounds are sometimes indistinguishable from unintentional errors unless they are self-reported as such, which is often unlikely due to perceived (or real) conflict with organizational policies.

In the high-risk industries, procedures are typically an aggregation of *steps* that provide instructions to the worker on specific sequence of *tasks* to accomplish an objective or goal. However, it is likely that some tasks, and some steps within tasks, are more complex than others. It is conceivable that deviations could be exacerbated for complex tasks or steps, where the worker may not be completely aware of the ramifications of a workaround or deviation because of inadequate or unclear information present in the procedure itself (Mearns, 2017; Park et al., 2003). Recent efforts have been made to quantify the task complexity stipulated in emergency operating procedures in nuclear industry (Park and Jung, 2007; Park et al., 2001) with reports of significant negative correlation between task complexity scores and task performance (Park et al., 2003). An important finding of Park et al. (2003) was that the complexity at the step-level may be a viable measure to predict the deviation behavior.

There is an abundance of published work on task complexity—identifying both that there are tasks which are inherently complex and that procedures steps can be written in a manner that creates complexity. We refer the reader to the synthesis of models, definitions, attributes, dimensions, and effects on performance in different industries and disciplines (Campbell, 1988; Liu and Li, 2012). Others have also explored the relationship between subjective and objective task complexity and how the perception of complexity impacts task performance (Maynard and Hakel, 1997). Despite numerous attempts to define task complexity, a *practical* method to measure the complexity of procedural tasks remains largely absent. Most previous efforts to articulate complexity require extensive contextual (e.g., tasks goals, location, event, etc.) and temporal information which in turn requires a significant amount of analytical resources (Campbell, 1988; Liu and Li, 2012; Maynard and Hakel, 1997). Given the large number of procedures used in even small petrochemical facilities, there is a crucial need for assessment techniques that are not only objective but also more efficient and potentially automated. Such methods will inform the design of new procedures and evaluation of existing procedures that allow for safer operations and provide metrics to assess cause-effect relationships among task complexity, workers' performance, and procedural adherence. To address this gap, we propose a view of task complexity grounded in the literature, present several measurable attributes to such complexity, and share a novel methodology (Complexity Index for Procedures—Step-level: CIPS) and proof of concept to quantify task-level complexity.

Our ultimate goal is to provide a methodology that can be applied for two related and distinct purposes. First, support procedure writers as they are creating procedures by providing information regarding complexity that the writers can use to make decisions regarding the wording in procedures. Second, given the vast number of existing procedures, provide an efficient and effective method for evaluating procedure complexity to identify those in need of revision. The CIPS is the first, but a necessary step toward that goal.

2. Method

2.1. Basis of the analysis

We established a basis for complexity by leveraging the concepts from Campbell (1988) and Park and Jung (2007). Campbell (1988) synthesized various attributes of task complexity across several research domains and proposed that task complexity is a function of (a) psychological experience, (b) task-person interaction, and (c) objective characteristics of the tasks, which is the focus of this research. According to Campbell, objective task characteristics contribute to task complexity based on increases in information load, information diversity, or rate of information change. Campbell's framework for the objective task complexity had four basic sources: multiple paths to a desired end state, multiple desired end states, conflicting interdependence, and uncertain or probabilistic linkages. Four categories of tasks were then operationalized based on the scaling of different sources: decision, problem, judgment, and fuzzy (Table 1).

Park and Jung (2007, 2008) developed and evaluated the task complexity measure which uses five sub-measures covering the quantifiable factors affecting the task complexity: step information, step logic, step size, abstraction hierarchy, and engineering decision (Table 2).

2.2. Method of quantifying complexity using natural language processing

Despite the promise shown by both task complexity measurement approaches, use of these frameworks on an industrial level requires a significant amount of context and time-consuming analysis. For example, measurement of sources in a problem or a fuzzy task in Campbell framework (1988) requires a detailed analysis of various goals, means to achieve each goal, as well as interdependence between subtasks. Similarly, modeling an abstraction hierarchy as in Park and Jung (2007) may contribute to better understanding of the work system, hence, the development of better procedures, however, this method requires deep knowledge of the system as well as specialized training. The application of these two concepts to thousands of different procedures such as would be required in high-risk industries such as chemical plants may prove to be a complex, tedious, and expensive endeavor. Hence, we integrated the underlying principles of these two frameworks to develop a simplified framework that can be applied to written procedures and automated using lexical analysis or natural language processing to produce a task complexity score.

One of the subtle similarities between Campbell (1988) and Park and Jung (2007) is the need to examine attributes of the steps of the task. Park and Jung (2007) focused explicitly on articulating step-level complexity and Campbell (1988) described tasks as a set of detailed steps. Thus, we chose to have step-level complexity as our focus of analysis. Grounded in Campbell framework, the proposed step-level

Table 1
Types of tasks and their associated sources of complexity adopted from Campbell (1988).

Task type	Complexity sources
Decision	Number of desired outcomes to attain; conflicting interdependence among outcome; uncertainty
Judgment	Conflicting and probabilistic nature of task information; uncertainty
Problem	Path multiplicity to a single desired outcome; conflicting interdependence among paths; uncertainty
Fuzzy	Outcome multiplicity; path multiplicity; conflicting interdependence; uncertainty

Table 2
Measures of task complexity and their associated complexity sources adopted from Park and Jung (2007).

Complexity measure	Complexity sources
Step information	Amount of information to be managed for the task
Step logic	Degree of logical entanglement due to the logical sequence of the required task
Step size	Number of required actions to accomplish the task
Abstraction hierarchy	Amount of system knowledge in recognizing the problem space
Engineering decision	Amount of cognitive resource in establishing appropriate decision criterion

complexity score uses five quantifiable characteristics of written procedures: (1) presence of multiplicity either in path or goal resulted in a decision making, (2) uncertainty or ambiguity or incompleteness present in the information that affects the performance of the step, (3) presence of interdependency among various instructions or steps, (4) multiplicity of instruction or condition or object of instruction in a particular step, and (5) amount of information need to be processed by the worker.

These characteristics can be used to assess the complexity across five attributes: *decision complexity*, *judgement complexity*, *interdependence*, *step size*, and *step information*. If the worker must make a *decision* between multiple courses of action based on a defined condition at a certain procedure step, this can be detected linguistically using various identifiers such as "if … then" or "either … or." Similarly, ambiguities or lack of clarity on information regarding the object of instruction may impose a judgment complexity where workers need to apply their engineering knowledge or *judgement* to accomplish the task. Such ambiguities can be detected linguistically, for example if the qualifier for an action verb is missing (e.g., "reduce flow" with "by 20%" missing). Dependency among more than one instruction (which may occur in the same step or different steps) constitute the *interdependence* complexity and can be detected (e.g., "Flush the pipe *before* opening the valve"). If a worker must complete more than one instruction in one step, this may impose *step size* complexity, detected linguistically when multiple action verbs or clauses (comma-separated or bullet points) are used, or when identifiers such as "and," "then," or "also" are used (e.g., "Set the flow to X, then document hourly values in the worksheet). If additional non-instructive or clarification information is included (e.g., a note to assist the worker in proper interpretation of the instruction), this may impose the *step information* complexity (See Table 3 for sample logics and identifiers for the five task complexity attributes).

2.3. Procedure complexity Index—Step-level

We propose a simple but practical approach to calculate a complexity index for procedures to quantify the presence of various attributes that contribute to step-level complexity linguistically. In particular, each step in a procedure can be parsed linguistically for the evidence of each attribute or source of step-level complexity. A step may contain more than one sources of complexity based on the nature of the task. The count of all the evidence will result in a score for each attribute. The total complexity score for the procedure can be calculated as the sum of all complexity attributes:

$$Sum\ \ of\ \ complexity\ \ sources, C_{Total} = \sum_{i=1}^{i=5} S_i$$

Where S_1 is the *decision* complexity, S_2 is the *judgement* complexity, S_3 is the *interdependence* complexity, S_4 is the *step size* complexity, and S_5 is the *step information* complexity. To determine the presence of the complexity in the procedure, some attributes of each source of complexity are defined and some identifiers that represent these attributes are established. Table 3 presents a selection of such attributes and identifiers for each type of complexity. These identifiers include syllables, key words, or phrases that would indicate a certain need from the user, which would add to the complexity of performing the task. This technique uses linguistic search which allows for algorithmic processing of the thousands of procedures. Table 4 demonstrates the application of the aforementioned methodology to measure complexity of steps from taken different procedures. The representative steps were taken from tasks from upstream and downstream processing facilities such as hydrogen containment analysis, gas purification, forklift docking and reboiler operation.

Table 3
Attributes and identifiers to represent different types of complexity.

Attribute	Sample Logic	Sample Identifiers
Decision complexity		
Operator is given a choice	If (condition) … then (action verb) While/when (condition) … then (action verb) As (condition) required/desired/accepted/permitted … (action verb) Either (action verb) … or (action verb)	Action verb: any Condition: any
Judgment complexity		
Qualifier of the action verb missing	(Action verb) (object) (missing/incomplete/ambiguous qualifier)	Action verb: raise, reduce, lower, minimize, hold Object: flow, temp, valve Complete qualifier: to full, at of rate of #, by #%, each
Condition of action missing	(Action verb) (state)	Action verb: establish, maintain, ensure, make sure, allow State: steady state, high, rise
Assumption that worker has the skillset or knowledge to do the action	(Action verb) (object)	Action verb: Verify, inspect, isolate, lockout, build permit
Lack of specificity	(Action verb) (object) (qualifier)	Qualifier: slowly, until, higher, lower, above, sufficient, too low, no longer, all
Interdependency		
Having step that cannot be skipped or completed out of order	Complete (Action verb) before (action verb) Before proceeding to next step …	Action verb: any
Having instruction to skip steps	(Action verb), then … (action verb)	Go to step #; proceed to step #
Step size		
Multiple instructions/objects/conditions	(Action verb) and (action verb) (object)	Action verb: any
Presence of comma separated objects; bulleted items	(Action verb) (object), (object), and (object)	Object: any
Step info		
Non-instructional notes; hazard statements.	Warning	Note:

3. Results and application

To leverage the step-level complexity index and to show proof of concept, we developed a novel natural language processing- (NLP) based algorithm (Ade et al., 2019) that identifies and highlights the elements in a procedure that contribute to complexity to inform the procedure writing task (see Fig. 1). It analyzes the procedure's text and identifies the words and phrases associated with potential increased task-level complexity.

The described algorithm was developed using the programming

Table 4
Measurement of step-level complexity for different steps taken from procedures in different facilities, i.e., upstream and downstream processing facilities such as hydrogen containment analysis, gas purification, forklift docking and reboiler operation.

Step Description	Complexity Attribute					Score
	Decision Task	Judgement Task	Interdependence	Step Size	Step Info	
If hydrogen concentration is greater than 0.2% and less than or equal to 0.5%, then perform the following steps. Otherwise, enter N/A.	*If-then*	-	-	*Multiple conditions*	-	2
Visually check that the "Back-up Venting Fan" is running.	-	*Check; Venting fan identifier missing*	-	-	-	2
Ensure each FC purifier is still isolated from the process and regeneration lines.	-	*Ensure; Each*	-	*Multiple objects*	-	3
On the RB column: A. Ramp RB column bottoms to 45% over 90 min B. Ramp the RB reflux controller to 14 Mg/h at a ramp rate of 4 Mg/h. C. Ensure that the RB sample system process lines are left on-line to prevent plugging during shutdown. *Note: Plugging can cause severe damage to the lines*	-	*Ensure*	-	*Multiple instructions*	*Note*	3
Check that the water pressure is greater than 1 bar.	-	*Check*	-	-	-	1
Locate the table labelled "Forklift Inspection Checklist".	-	-	-	-	-	0
Once the trailer is docked, enter the building to activate the dock lock light system.	-	-	*Once*	*Multiple instructions*	-	1
When the SM Loop <20 ppm moisture, then line-up the SM through either A or B SM purifier to the SM recycle drum.	*When-then; Either-or*	-	-	-	-	2
Go to the "Back-up Venting Fan" and inspect by clicking on the Fan.	-	*Inspect*	-	*Multiple instructions*	-	2
Adjust the steam bypass valves as required to maintain reboiler or reboiler KO pot levels.	*As required*	*Adjust; Maintain*	-	*Multiple objects (or)*	-	4

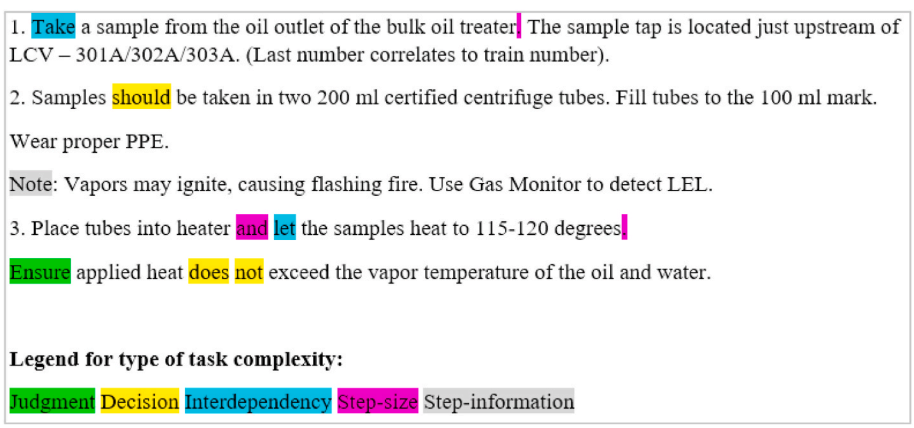

Fig. 1. A sample screenshot for a tool that highlights task-level complexity indicators for the procedure writer.

language Python 3.6 (Python Software Foundation, 2016) and the Natural Language Toolkit (NLTK) source library (Loper and Bird, 2002). The algorithm relies on a base vocabulary consisting of linguistic identifiers for the 5 types of task complexity. Apart from a simple text matching of linguistic identifiers, the algorithm utilizes the unsupervised machine learning algorithm GloVe (Pennington et al., 2014) to identify words and phrases that are semantically similar to the linguistic identifiers in the base vocabulary. The GloVe model allows vector representation of words to help quantify similarity between two or more words. The pre-trained 'GloVe6B-200 d model' (Pennington et al., 2014) has a vocabulary of 400,000 words that are used more frequently wherein each word is represented as a 200-dimensional vector. While applying the algorithm for a procedure, each word within each step of the procedure is compared with the linguistic identifiers along with an evaluation of semantic similarity to linguistic identifiers. Using this comparison, the algorithm identifies both the word and its associated type of task complexity. New word or phrases identified through similarity comparison are stored back in the base vocabulary, thus supporting the learning nature of the algorithm by improving its memory with respect to linguistic identifiers.

The described algorithm was tested on a set of 20 procedures (including a total of 550 steps) that included various tasks in the oil and gas and chemical industry (Ade et al., 2019). These procedures were obtained from multiple stakeholders in the Next Generation Advanced Procedures consortium. To quantify the accuracy of the algorithm, the words identified through the algorithm were stored in the vocabulary of the algorithm following its application. The resulting vocabulary contained false positives that were not necessarily associated with the specific task complexity. The approximate accuracies of the algorithm for different types of task complexities are shown in Table 5. It is important to note that the learning behavior of the NLP algorithm is

Table 5
Accuracy of the NLP algorithm for different types of task complexity.

Task complexity	Accuracy (%)
Judgment	79.3
Decision	60.0
Interdependency	60.0
Step-size	73.6
Step-information	71.4

controlled by the critical values of semantic similarity. Lower critical values result in faster and improper/inaccurate learning of new identifiers whereas, higher values inhibit the process of learning completely. Adjusting the critical values for each type of task complexity could result in an overall higher accuracy for the algorithm.

As described previously, the described framework identifies 5 attributes of step-level complexities. The identified complexities can possibly be reduced/eliminated by redesigning the procedures through adjustment of language, numbering, and added specifics. The redesigned procedure with reduced task complexities has the potential to improve the performance and safety of high-risk industries (McDonald et al., 2020). Examples of such procedure step redesign are depicted in Table 6. The depicted examples are obtained from the testing set of 20 procedures.

Table 6
Redesigning of procedure steps from a complexity perspective.

Step description	Task complexity	Redesigned step
Adjust operating parameters to maintain mainline rate and pressure	Judgment (maintain)	Adjust the operating parameters to maintain rate and pressure at 20 bbl/min and 120 bar
a) If fan belt is acceptable, proceed to step c. b) If fan belt is unacceptable, proceed to step d. c) Open safety guards d) Perform LOTO procedure and readjust the belt.	Interdependency (multiple sub-steps)	If the belt is acceptable, open safety guards. If the belt is unacceptable, perform LOTO procedure and readjust the belt.
Increase test pressure until PSH is tripped	Decision (until)	Increase test pressure above 120 psi to achieve tripping of PSH
Open block valve downstream of the control valve CV101 and close gas inlet valve upstream of the control valve CV101	Step-size (and)	a) Open block valve downstream of the control valve CV101 b) Close gas inlet valve upstream of the control valve CV101

4. Discussion and conclusions

In this paper, we presented a novel complexity measurement framework that uses five objective attributes of step-level complexities that were grounded in existing task complexity evaluation theories. The framework and the associated complexity index provide a practical method for the design and assessment of written procedures. This method can be used across several high-risk industries in which written procedures are prevalent. In conjunction with the process hazard information, complexity quantification provides vital information to the procedure writers and enables an efficient procedure management cycle which results in task information presentation for the safe completion of the operations across several high-risk industries in which written procedures are prevalent.

While this paper demonstrates the efficacy of the step-level complexity index through a case study, the presented identifiers or linguistic search logics are by no means collectively exhaustive. Indeed, the presented methodology is at its infancy and there is a need for rigorous verification and validation. Several important limitations are noteworthy. First, the proposed methodology relies heavily on linguistics parsing methods. Given the relative ease of generating this metric and automated scaling up to encompass all procedures in a facility, a retrospective analysis with available safety data of a facility, can provide a starting point to procedure writers. Second, the methodology uses simple summation of counts of evidence for each attribute. More work is needed to understand the impact of each attribute on performance to inform a weighted summation approach. Finally, while the measurement of step-level complexity index using NLP methods shows promise, it is important for such method to complement the procedure writing process as a hybrid system involving human judgment because reduction of complexity does not necessarily improve the quality of the procedure. For example, elimination of a note from a step reduces the *step information* complexity but may make the step ambiguous and increase the *judgement* complexity. Therefore, human interpretation is needed to determine when one complexity may be appropriate to make the step clear or comprehensive.

Despite these limitations, the complexity evaluation framework presented here may allow organizations to evaluate potential issues present in their current procedures and avoid unnecessary complexity in new procedures, by quantifying and reducing complexity objectively. This may in turn reduce the perceived complexity and improve compliance/deviation behavior which may lead to an improved safety culture and efficiency. In addition, the current lack of performance data associated with other types of complexity provides future opportunity to perform controlled experiments to identify correlations and develop a robust framework. More importantly such framework also incorporates a use-review-learn-improve cycle allowing organizations to modify the attribute list based on local variations. Future work also includes the deployment of the other complexity parameters either using controlled laboratory experiments in pilot industry environments or theoretical modeling to objectively evaluate the interactions between the procedures, environments, and the human-element, which ultimately offers the potential for safer operations in these high-risk environments.

Funding

This work was supported by the Texas A&M University's Mary Kay O'Connor Process Safety Center and NGAP (the Next Generation Advanced Procedures Initiative) at Texas A&M University.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The researchers would like to thank the industry collaborators for the access to procedures utilized for the development and testing of the complexity concept.

References

Ahmed, L., Quddus, N., Kannan, P., Peres, S.C., Mannan, M.S., 2020. Development of a procedure writers' guide framework: integrating the procedure life cycle and reflecting on current industry practices. Int. J. Ind. Ergon. 76 https://doi.org/10.1016/j.ergon.2020.102930.

Ade, N., Quddus, N., Kannan, P., Peres, S.C., Parker, T., 2019. Application of natural language processing in procedure design. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 63 (1), 257–261. https://doi.org/10.1177/1071181319631108.

Bullemer, P.T., Hajdukiewicz, J.R., 2004. A study of effective procedural practices in refining and chemical operations. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 48 (20), 2401–2405. https://doi.org/10.1177/154193120404802006.

Campbell, D.J., 1988. Task complexity: a review and analysis. Acad. Manag. Rev. 13 (1), 40–52. https://doi.org/10.5465/amr.1988.4306775.

Chemical Safety and Hazard Investigation Board, 2017. ExxonMobil Torrance Refinery Electrostatic Precipitator Explosion, pp. 1–73.

Chemical Safety and Hazard Investigation Board, 2018. Completed investigations. Retrieved from. https://www.csb.gov/investigations/completed-investigations/?F_FromDate=1997&F_ToDate=2000.

Degani, A., Wiener, E.L., 1997. Procedures in complex systems: the airline cockpit. IEEE Trans. Syst. Man Cybern. Syst. Hum. 27 (3), 302–312. https://doi.org/10.1109/3468.568739.

Holmstrom, D., Altamirano, F., Banks, J., Joseph, G., Kaszniak, M., MacKenzie, C., Wallace, S., 2006. CSB investigation of the explosions and fire at the BP Texas City refinery on March 23, 2005. AIChE Annual Meeting. Conf. Proc.

Lattanzio, D., Patankar, K., Kanki, B.G., 2008. Procedural error in maintenance: a review of research and methods. Int. J. Aviat. Psychol. 18 (1), 17–29. https://doi.org/10.1080/10508410701749381.

Liu, P., Li, Z., 2012. Task complexity: a review and conceptualization framework. Int. J. Ind. Ergon. 42 (6), 553–568. https://doi.org/10.1016/j.ergon.2012.09.001.

Loper, E., Bird, S., 2002. NLTK: the natural language Toolkit. Retrieved from. http://arxiv.org/abs/cs/0205028.

Maynard, D.C., Hakel, M.D., 1997. Effects of objective and subjective task complexity on performance. Hum. Perform. 10 (4), 303–330. https://doi.org/10.1207/s15327043hup1004_1.

McDonald, A.D., Ade, N., Peres, S.C., 2020. Predicting procedure step performance from operator and text features: a critical first step toward machine learning-driven procedure design. Human factors. https://doi.org/10.1177/0018720820958588.

Mearns, K., 2017. Human factors in the chemical process industries, pp. 149–200. https://doi.org/10.1016/bs.mcps.2017.01.002.

Park, J., Jung, W., 2007. A study on the development of a task complexity measure for emergency operating procedures of nuclear power plants. Reliab. Eng. Syst. Saf. 92 (8), 1102–1116. https://doi.org/10.1016/j.ress.2006.03.009.

Park, J., Jung, W., 2008. A study on the validity of a task complexity measure for emergency operating procedures of nuclear power plants-Comparing task complexity scores with two sets of operator response time data obtained under a simulated SGTR. Reliab. Eng. Syst. Saf. 93 (4), 557–566. https://doi.org/10.1016/j.ress.2007.02.002.

Park, J., Jung, W., Kim, J., Ha, J., 2003. The step complexity measure - its meaning and applications. Journal of Korean Nuclear Society 35 (1), 80–90.

Park, J., Jung, W., Kim, J., Ha, J., Shin, Y., 2001. The step complexity measure for emergency operating procedures - comparing with simulation data. Reliab. Eng. Syst. Saf. 74 (1), 63–74. https://doi.org/10.1016/S0951-8320(01)00063-1.

Pennington, J., Socher, R., Manning, C., 2014. Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543. https://doi.org/10.3115/v1/D14-1162.

Peres, S.C., Quddus, N., Kannan, P., Ahmed, L., Ritchey, P., Johnson, W., Mannan, M.S., 2016. A summary and synthesis of procedural regulations and standards—informing a procedures writer's guide. J. Loss Prev. Process. Ind. 44, 726–734. https://doi.org/10.1016/j.jlp.2016.08.003.

Python Software Foundation, 2016. Python Software Foundation. Python Language Reference, Version 3.6.

Williams, J.P., Sasangohar, F., Peres, S.C., Smith, A., Mannan, M.S., 2017. Investigating written procedures in process safety: qualitative data analysis of interviews from high risk facilities. Proceedings of the Human Factors and Ergonomics Society 1669–1670. https://doi.org/10.1177/1541931213601905, 2017-Octob.