Nuclear Engineering and Technology 54 (2022) 2037–2047

Contents lists available at ScienceDirect

Nuclear Engineering and Technology

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

Original Article

# Empirical estimation of human error probabilities based on the complexity of proceduralized tasks in an analog environment

Jinkyun Park*, Hee Eun Kim, Inseok Jang

*Korea Atomic Energy Research Institute (KAERI), South Korea*

ARTICLE INFO

*Article history:*
Received 21 July 2021
Received in revised form
22 November 2021
Accepted 20 December 2021
Available online 27 December 2021

*Keywords:*
Task complexity
Nuclear power plant
Analog environment
Proceduralized task
Error of omission
Error of commission
Logistic regression

ABSTRACT

The contribution of degraded human performance (e.g., human errors) is significant for the safety of diverse social-technical systems. Therefore, it is crucial to understand when and why the performance of human operators could be degraded. In this study, the occurrence probability of human errors was empirically estimated based on the complexity of proceduralized tasks. To this end, Logistic regression analysis was conducted to correlate TACOM (Task Complexity) scores with human errors collected from the full-scope training simulator of nuclear power plants equipped with analog devices (analog environment). As a result, it was observed that the occurrence probability of both errors of commission and errors of omission can be soundly estimated by TACOM scores. Since the effect of diverse performance influencing factors on the occurrence probabilities of human errors could be soundly distinguished by TACOM scores, it is also expected that TACOM scores can be used as a tool to explain when and why the performance of human operators starts to be degraded.


## 1. Introduction

According to Baxter and Sommerville [1], the term *socio-technical systems* was initially suggested by Emery and Trist [2] to indicate "systems that involve a complex interaction between humans, machines and the environmental aspects of the work system". A more popular definition of socio-technical systems is also available: "A system involving the interaction of hard systems and human beings, in ways that either cannot be separated or are thought to be inappropriate to separate" [3]. From these definitions, it is strongly anticipated that the safe operation of complicated process control systems included in diverse industrial sectors (e.g., chemical plants, transportation systems, maritime vessels, railway systems, aviation systems) is largely dependent on both the reliability of hardware and the reliability of human operators. Indeed, including nuclear power plants (NPPs), this is why the degradation of human performance (e.g., human errors) is attributable as the root cause of significant accidents experienced in the last several decades [4–7].

Therefore, a huge amount of effort has been spent to prevent the degradation of human performance, including the identification of key factors affecting the performance of human operators (performance influencing factors, PIFs, or performance shaping factors, PSFs) [8]. In this regard, Park [9] summarized seven groups of PIFs detailing those that are meaningful for understanding the variation in the performance of human operators, which were found from existing literature. The seven PIF groups are as follows: (1) Operator characteristics, (2) Social aspects, (3) Task, (4) Environment, (5) Organization, (6) System, and (7) Human–machine interface system (HMIS).

Each PIF group contains one or more detailed PIFs. For example, the *Environment* PIF group includes detailed PIFs such as 'temperature', 'humidity', 'noise', and 'vibration,', while the *System* PIF group encompasses 'rate of change of critical parameters', 'number of changing variables', and 'highly unstable plant situation'. In addition, in the case of the *Task* PIF group, there are many detailed PIFs that can be subdivided into two sub-categories such as *Task contents* and *Task type/attribute*.

- Task contents: (1) Procedure type, (2) Procedure availability, (3) Amount of required information, (4) Logic structure, (5) Decision-making criteria, (6) Clarity of instructions and terminologies

* Corresponding author. 989-111, Daedeok-daero, Yuseong-gu, Daejeon, 34057, South Korea.
*E-mail address:* kshpjk@kaeri.re.kr (J. Park).

https://doi.org/10.1016/j.net.2021.12.025

- Task type/attribute: (1) Type of cognitive activities (e.g., monitoring and detection), (2) Required level of cognition, and (3) Dynamic/step-by-step task

Accordingly, if the effect of each detailed PIF on the performance of human operators can be soundly correlated, it is expected that their performance variation in a specific situation can be properly estimated by the combination of the abovementioned seven PIF groups. Along these lines, Park [10] proposed a measure, TACOM (Task Complexity), that can quantify the level of task complexities to be loaded by human operators who have to accomplish a series of proceduralized tasks. Park [9] claimed that TACOM could be regarded as a normative task complexity measure because it can cover the significant PIFs belonging to the *Task* PIF group. If so, it is expected that there should be a significant relation between the performance of human operators and the associated TACOM scores.

Indeed, Podofillini et al. [11], pointed out that the likelihood of human error observed from the full-scope training simulator of an NPP seems to be closely related to TACOM scores. In addition, Park [12] observed that not only the subjective workloads but also the response times of human operators are proportional to an increase of TACOM scores. More recently, Jang et al. [13] revealed that there is a notable correlation between TACOM scores and the number of human errors observed from simulated emergency situations in a fully-digitalized main control room (MCR). Such results of previous studies support that the TACOM measure could play a significant role in the estimation of the occurrence probability of human errors.

In this study, the occurrence probability of human error is empirically investigated with respect to TACOM scores in detail. To this end, an inventory of human error data collected from a full-scope training simulator of a Korean domestic NPP is revisited; the simulator is a replica of an analog MCR equipped in a Westinghouse 3-loop pressurized water reactor plant. As a result, it was observed that there is a statistically meaningful correlation between the occurrence probability of human errors and the associated TACOM scores. Specifically, compared to errors of omission (EOOs), the occurrence probability of errors of commission (EOCs) showed a clearer correlation with TACOM scores.

The rest of this paper are structured as follows. First, in Section 2, key features of the TACOM measure are introduced based on a review of existing studies. After that, raw information of human errors obtained from the analog MCR environment is explained in Section 3. Then in Section 4, detailed descriptions are provided about how to empirically estimate the occurrence probabilities of both EOCs and EOOs by using the logistic regression technique. Finally, in Section 5, the conclusion of this study is drawn after discussing limitations and future work to be followed.

## 2. Key features of the TACOM measure

### 2.1. Overview of TACOM measure

As briefly explained in Section 1, it is important to manifest when and why the performance of human operators could be degraded. This implies that the very first phase is to understand their performance variations with respect to the status of many PIFs belonging to at least one of the seven groups. In this regard, it is essential to consider the characteristics of the tasks described in related operation procedures. This is because the operational experience of diverse industrial sectors has emphasized that the use of procedures is one of the strongest countermeasures against the degradation of human performance [14–16].

Without loss of generality, the contents of a procedure such as an emergency operating procedure (EOP) in NPPs can be subdivided into proceduralized tasks, procedural steps, and detailed actions that instruct what should be done and how to conduct it. Fig. 1 depicts an overall structure of a procedure with proceduralized tasks and associated procedural steps. At a glance, the preparation of a well-written procedure with detailed procedural steps appears to be sufficient for preventing the degradation of human performance.

Following a procedure is harder than it seems, however, because human operators have to deal with drastically changing situations (dynamic situations) by using a series of predefined rules (static prescriptions) that cannot reflect the whole spectrum of such dynamic situations [17,18]. In other words, even though the contents of a procedure can be regarded as a catalog of predefined rules, performing a procedure is not the following of a simple IF-THEN-ELSE rule but rather a kind of cognitive work composed of diverse activities including (but not limited to) (1) monitoring process parameters, (2) interpreting the nature of an on-going situation, (3) planning appropriate responses to properly cope with the situation at hand, and (4) implementing detailed actions prescribed in a procedure. This indicates that, in order to clarify the effect of the *Task* PIF group on the performance of human operators, it is crucial to evaluate the complexity level of a proceduralized task first. For this, the TACOM measure proposed in Ref. [10] can be used to quantify the level of proceduralized tasks to which human operators are exposed (Fig. 2).

In brief, the TACOM measure consists of five sub-measures that are able to numerically evaluate the effect of significant factors making the performance of proceduralized tasks complicated by using the first and second order of graph entropy concepts. More detailed information about the quantification of each sub-measure can be found in Ref. [10].

### 2.2. Appropriateness of TACOM measure

As can be seen from Fig. 2, once the contents of a proceduralized task are specified, it is necessary to conduct a task analysis to extract key information pertaining to its complexity. The extracted information is then used to calculate the values of five sub-measures via relevant mathematical concepts (i.e., the first- and second-order graph entropies). Fig. 3 schematically shows the calculation of three sub-measures, namely SIC (step information complexity), SLC (step logic complexity), and SSC (step size complexity).

For example, the second-order graph entropy is applied to obtain the SIC value that represents the complexity of a proceduralized task from the perspective of 'the amount of information to be processed by human operators'. In contrast, the first-order graph entropy allows us to assess the SSC value that represents the complexity of a proceduralized task due to its predefined action sequence to be followed by human operators. When all of the values for the five sub-measures are specified, the TACOM score of a proceduralized task is determined using the formula shown in the top of Fig. 2.

Following the development of the TACOM measure, its appropriateness should be systematically investigated. In this regard, from both qualitative and quantitative perspectives, it is important to clarify at least the following two technical issues: (1) the coverage of the TACOM measure in terms of the *Task* PIF group, and (2) the relationship between the performance of human operators and TACOM scores.

The first technical issue related to the coverage of the TACOM measure can be addressed by the following two separate questions: (1) Can the TACOM measure properly cover the detailed PIFs included in the *Task contents* sub-category? and (2) Can the TACOM measure be applied to explain the characteristics of the detailed

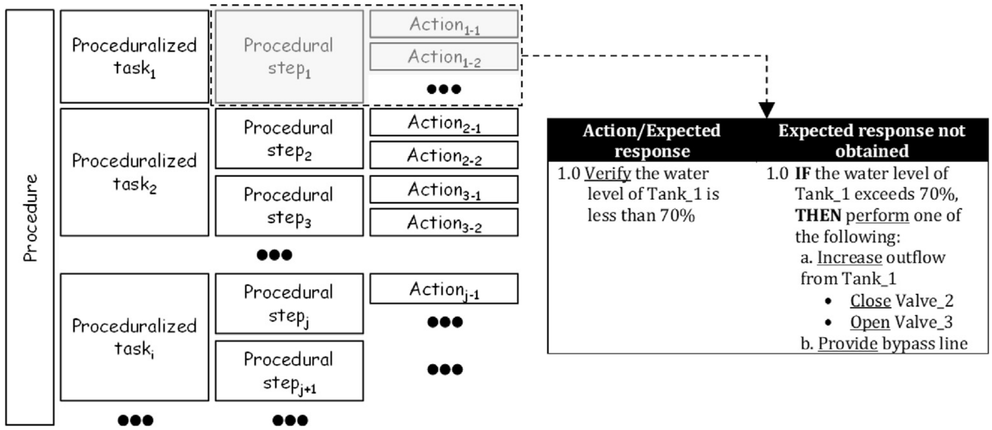



**Fig. 1.** Proceduralized tasks, procedural steps, and detailed actions; modified from [10].

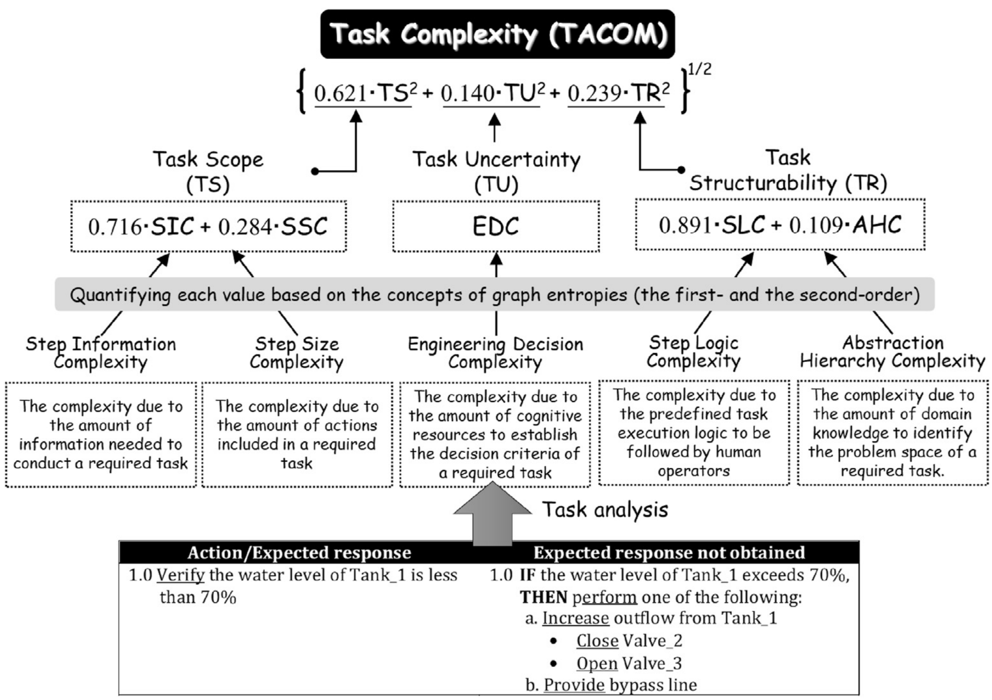



**Fig. 2.** Overall process to calculate the TACOM score of a proceduralized task; modified from [9].

PIFs belonging to the *Task type/attribute* sub-category? In order to answer the first question, it is necessary to investigate whether or not the TACOM measure can reflect the effects of the *Task contents* PIFs in quantifying the complexity level of a proceduralized task. Similarly, in terms of the second question, if the TACOM measure is irrelevant to distinguish the characteristics of task types (e.g., step-by-step task vs. dynamic task), it is hard to confirm its appropriateness. From this concern, Park [9] claimed that the TACOM measure is relevant because it not only contains detailed PIFs belonging to the *Task contents* sub-category but also represents the key characteristics of detailed PIFs related to the *Task type/attribute* sub-category.

The second technical issue is the validity of the TACOM measure. That is, if the TACOM measure has a sufficient coverage in quantifying the complexity level of a proceduralized task, it is natural to expect that TACOM scores are statistically meaningful for explaining the variability of human performance. In this regard, many researchers have pointed out that there are meaningful relationships between TACOM scores and the diverse dimensions of human performance, such as response times, subjective workloads, and human error rates [11,12,19–21]. A more interesting result reported by Jang et al. [13] indicates that the number of human errors observed from the full-scope training simulator of NPPs, which is a replica of a fully digitalized MCR, exponentially goes up with an increase in TACOM scores. If there is a notable relationship between the number of human errors and the associated TACOM scores, it can be confidently assumed that human error probability (HEP) can be empirically estimated as a function of TACOM score. In order to scrutinize this assumption, human performance data collected from the full-scope training simulator of NPPs, which is the replica of an analog MCR, are revisited in this study.

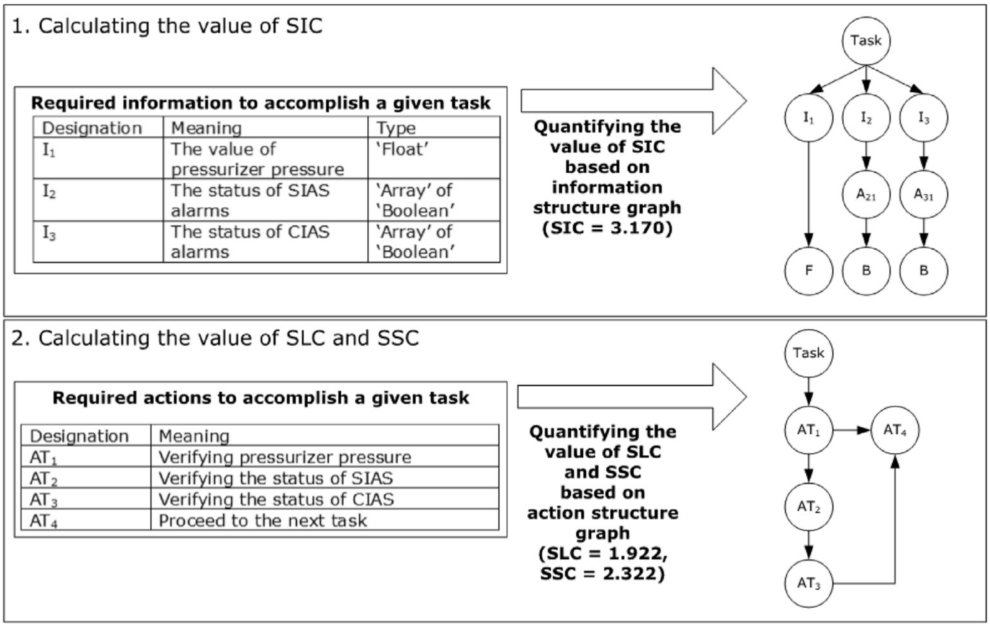



**Fig. 3.** Example of TACOM's sub-measure calculations; adopted from [11].

## 3. Human error data available from an analog environment

### 3.1. Revisiting human performance data

One of the representative approaches to investigate the effects of diverse PIFs on the performance of human operators is to precisely scrutinize their responses observed from an actual working environment. Unfortunately, in the case of studying the performance variation of human operators under incident or accident conditions, this approach is unrealistic because we are not able to intentionally initiate any events resulting in unwanted consequences. As an alternative, human performance data observed from simulated incident or accident conditions have been frequently analyzed for many decades [22]. In general, a partial- or full-scope simulator can be used for this purpose, and dedicated frameworks to facilitate the collection of human performance data are available now.

For instance, KAERI (Korea Atomic Energy Research Institute) developed a framework namely HuREX (Human Reliability data Extraction) that allows us to collect various kinds of human performance data (including human errors) from the full-scope training simulator of NPPs [23]. With this framework, KAERI developed a large archive called the HuREX database that contains human performance data from two kinds of full-scope training simulators established in domestic Korean NPPs [24–26]. That is, although most domestic Korean NPPs are controlled by analog MCRs equipped with traditional devices (e.g., chart recorders, alarm tiles, push buttons, and hand switches), there are several NPPs operated by fully digitalized MCRs installed with up-to-date devices (e.g., alarm display screens, information display screens, and soft controls). As these two working environments are significantly different, KAERI independently gathered two sets of human performance data. It should be noted that, in contrast to Jang et al. [13] who compared the number of human errors observed from the full-scope training simulator of a digitalized MCR with TACOM scores, the present study focuses on those of an analog MCR. Table 1 summarizes the catalog of accident scenarios with the number of simulations that were secured from the full-scope training simulators of analog MCRs with the participation of human operators who are actually working as MCR operators in NPPs.

As can be seen from Table 1, human performance data were collected from a large number of simulation records representing diverse abnormal and accident conditions. Of them, in terms of quantifying the complexity level of a proceduralized task as depicted in Fig. 1, simulation records obtained from abnormal conditions are inappropriate because human operators pick out the most relevant abnormal operating procedure (AOP) from the catalog of AOPs based on their own decision.

In other words, since there is no *dedicated* procedure for supporting how to select a proper AOP when an abnormal condition occurs, it is difficult for human operators to enter the AOP that fits the abnormal condition at hand [27,28]. Indeed, the excerpt below clearly explains this situation.

> "During abnormal situations, a well-trained operator should comprehend a malfunction in real time by analyzing alarms, assessing values, or recognizing unusual trends of multiple instruments [...]. In an NPP, many alarms from many different systems often occur at the same time during an incident, making it difficult for the operator to select a correct response efficiently. Too many information imposes a heavy burden on operators in a time-critical situation, and it is very difficult for them to conduct a thorough assessment of each individual symptom in a short period of time" [29; p. 413].

In contrast, when an accident occurs, human operators use a specific procedure that allows them to recognize its nature based on diverse symptoms (e.g., the status of a component and the trend of a key process parameter) [30,31]. This implies that it is possible to compare the complexity level of proceduralized tasks (i.e., TACOM scores) with the performance of human operators who are trying to minimize the consequence of the accident by using a procedure. Accordingly, in this study, human performance data gathered from accident conditions (refer to the first and second row of Table 1) were revisited to obtain the raw information of human errors observed from simulated accident conditions.

**Table 1**
Inventory of simulation records collected from the full-scope training simulators of analog MCRs.

| No. | Simulation scenario | Number[a] | Period[b] | Condition[c] |
|---|---|---|---|---|
| 1 | ISLOCA[d] | 10 | Sep. 2009 to Dec. 2009 | Accident |
| 2 | MSLB[e] followed by SGTR[f] | 8 | May 2010 to Aug. 2010 | |
| 3 | Slip down of control rods | 14 | May 2007 to Dec. 2008 | Abnormal |
| 4 | Wrong operation of VCT[g] outlet valve | 18 | | |
| 5 | Failure of pressurizer level controller | 22 | | |
| 6 | High vibration of containment cooling fan | 18 | | |
| 7 | Failure of deaerator inlet valve | 13 | | |
| 8 | Condenser tube leak | 40 | | |
| 9 | Loss of instrumentation air | 19 | | |
| 10 | Failure of emergency seal oil pump | 22 | | |
| 11 | Loss of non-essential electric power | 10 | | |
| 12 | Malfunction of cyclone filter | 8 | Sep. 2011 to Nov. 2011 | Abnormal |
| 13 | Decrease of deaerator water level | 8 | | |
| 14 | Loss of condenser vacuum | 13 | | |

[a] Number of simulations.
[b] Collection period.
[c] Simulation condition.
[d] Interfacing system loss of coolant accident.
[e] Main steam line break.
[f] Steam generator tube rupture.
[g] vol control tank.

### 3.2. Extracting raw information of human error

The HuREX framework defines a human error as “an action inappropriately taken by plant personnel, or not taken when needed, resulting in a degraded plant safety condition” [23]. From this definition, it is evident that the HuREX framework considers two types of human errors: EOC (an action inappropriately taken) and EOO (an action not taken when needed). Unfortunately, the identification of human errors is still vague unless *the degraded plant safety condition* is clearly stipulated. For this reason, the HuREX framework designates that “A human error is thus identified by the operator's responses meeting the following two criteria: (1) the response failed to satisfy the performance requirements stated in the procedures, and (2) the response has any of the following three consequences: (a) inappropriate procedure/step transition, (b) inappropriate equipment control, and (c) inappropriate communication to the MCR outside” [22]. On the basis of the HuREX framework, various kinds of raw information are accumulated in the HuREX database, which is helpful for understanding when and why human errors occurred under simulated accident conditions.

Of the available raw information in the HuREX database, in terms of estimating an empirical human error probability as the function of a TACOM score, it is necessary to extract at least the following raw information: (1) types of human error (EOO or EOC), (2) procedural steps in which human errors occurred, and (3) how many times the corresponding procedural steps were conducted by human operators. Once the abovementioned raw information is obtained, the TACOM score of each procedural step in which a human error was observed can be calculated. For example, in the case of the procedural step exemplified in Fig. 1, if a human operator read the water level of Tank_1 as 80% when its correct value was 60%, several EOOs can be counted because the actions for coping with the lowered water level were not properly conducted (e.g., Close Valve_2). Table 2 summarizes the abovementioned raw information with the associated TACOM scores identified from the HuREX database.

As can be seen from Table 2, human errors occurred during the performance of 19 procedural steps. In other words, most of the procedural steps in the simulated accident conditions were successfully conducted with no human errors. Indeed, this is a desirable tendency for human operators who are responsible for the operation of socio-technical systems such as NPPs. At the same time, however, this implies that Table 2 may be insufficient for the estimation of empirical HEPs. For this reason, a series of hypothetical proceduralized tasks were introduced in this study. Fig. 4 would be helpful for explaining this idea.

As depicted in Fig. 4, let us assume that two human errors were observed in *Procedural step$_1$* while one human error was observed in *Procedural step$_2$*. With these observations, since *Proceduralized task$_2$* consists of both *Procedural step$_2$* and *Procedural step$_3$*, it is possible to say that one human error occurred during the performance of *Proceduralized task$_2$*. In this case, the TACOM score of *Proceduralized task$_2$* can be quantified by the contents of both *Procedural step$_2$* and *Procedural step$_3$*. Similarly, it can be said that three human errors occurred during the performance of *Hypothetical proceduralized task$_1$* that is composed of both *Proceduralized task$_1$* and *Proceduralized task$_2$*. Therefore, if we calculate the TACOM score of *Hypothetical proceduralized task$_1$*, the amount of data included in Table 2 can be soundly expanded.

In generating a hypothetical proceduralized task, two rules were considered. The first one is that the procedural steps of which the value of human error fraction given in the last column of Table 2 is too high should be discarded from the generation of a hypothetical proceduralized task. For example, let us recall ‘Verify the water level of Tank_1 is less than 70%’, one of the detailed actions depicted in Fig. 1. If a human operator wrongly reads the actual water level of Tank_1 (e.g., 50%), all of the detailed actions that are supposed to be conducted when its water level is less than 70% (e.g., ‘Close Valve_2’ and ‘Open Valve_3’) will be entirely omitted. In the HuREX framework, since each omission should be counted as the total number of EOOs pertaining to the performance of reading the actual water level of Tank_1, there are times when the number of total human errors approaches the number of trials (refer to the 16th, 18th, and 19th procedural steps in Table 2) or even exceeds it (refer to the 14th, 15th, and 17th procedural steps in Table 2). In other words, there are times when single ‘Action’ included in a procedural step results in two or more EOOs. Accordingly, these procedural steps were discarded from the generation of hypothetical proceduralized tasks because they are able to distort the occurrence number of human errors with respect to TACOM scores.

The second rule is that the scope of a hypothetical proceduralized task should be restricted to the procedural steps that are successive. For example, it is irrelevant to define a hypothetical

**Table 2**
TACOM scores and human errors observed during the performance of individual procedural steps.

| ID | Trial[a] | Number of EOOs | Number of EOCs | TACOM score | Human error fraction[b] | Remark |
|---|---|---|---|---|---|---|
| 1 | 17 | 1 | 2 | 4.084 | 0.18 | – |
| 2 | 18 | 3 | 0 | 3.813 | 0.17 | – |
| 3 | 18 | 2 | 0 | 2.904 | 0.11 | – |
| 4 | 10 | 1 | 0 | 3.050 | 0.10 | – |
| 5 | 9 | 3 | 0 | 2.806 | 0.33 | – |
| 6 | 9 | 3 | 0 | 2.755 | 0.33 | – |
| 7 | 9 | 1 | 1 | 3.424 | 0.22 | – |
| 8 | 9 | 2 | 0 | 3.436 | 0.22 | – |
| 9 | 9 | 2 | 2 | 5.151 | 0.44 | – |
| 10 | 8 | 1 | 0 | 2.806 | 0.13 | – |
| 11 | 8 | 0 | 1 | 3.832 | 0.13 | – |
| 12 | 7 | 0 | 4 | 3.350 | 0.57 | – |
| 13 | 6 | 1 | 0 | 2.896 | 0.17 | – |
| 14 | 9 | 3 | 9 | 3.027 | 1.33 | Discard |
| 15 | 8 | 10 | 0 | 4.680 | 1.25 | Discard |
| 16 | 7 | 0 | 6 | 2.845 | 0.86 | Discard |
| 17 | 7 | 18 | 1 | 3.179 | 2.71 | Discard |
| 18 | 6 | 5 | 0 | 5.158 | 0.83 | Discard |
| 19 | 1 | 0 | 1 | 3.337 | 1.00 | Discard |

[a] The total number of executions with respect to each procedural step.
[b] (Number of EOOs + Number of EOCs)/Trial.

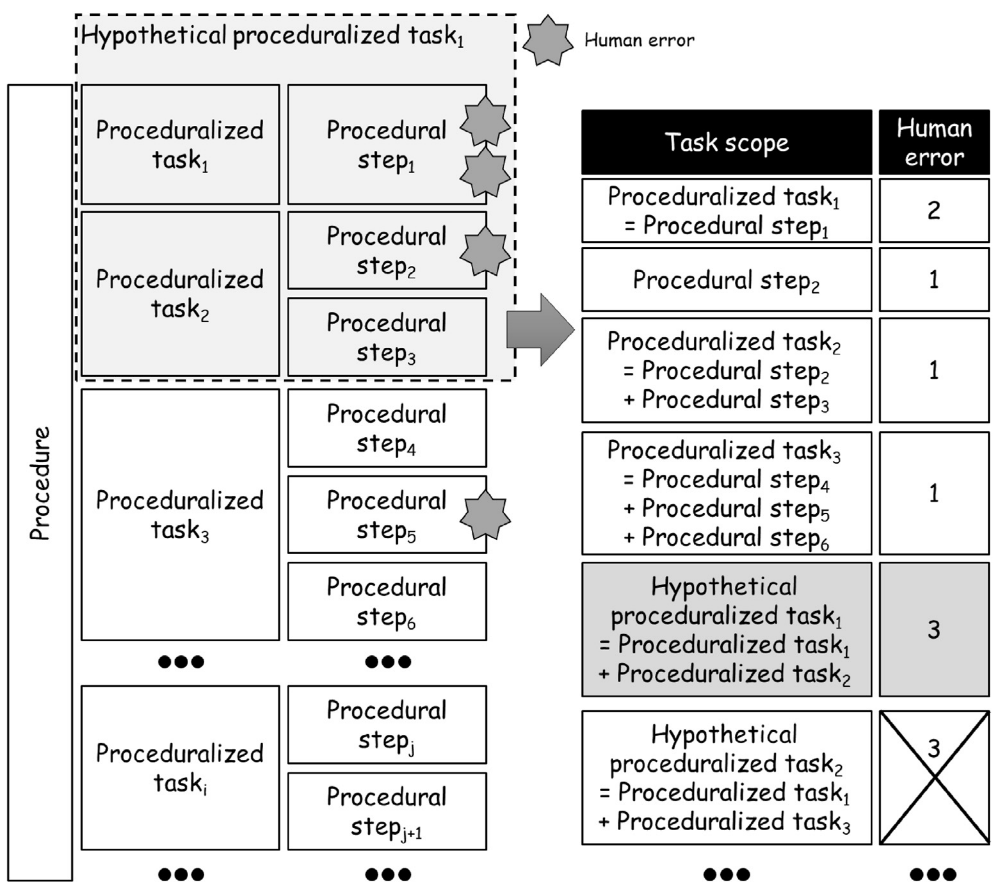



**Fig. 4.** Example of counting human errors with respect to hypothetical tasks; modified from [32].

proceduralized task that consists of both *Procedural step$_1$* and *Procedural step$_3$*. In other words, without the consideration of *Procedural step$_2$*, this corresponds to the null definition of a proceduralized task that does not have an actual meaning. For example, as shown in Fig. 4, if we assume a hypothetical task that consists of two subtasks (i.e., ‘Proceduralized task$_1$’ and ‘Proceduralized task$_3$’), the number of data points available will be easily increased. However, in the case of generating additional data points, the physical meaning of a merged task should be carefully clarified. In this light, the abovementioned example is irrelevant because there is no rationale justifying the physical meaning of the merged task. That is, even though we create a new task by combining ‘Proceduralized task$_1$’ and ‘Proceduralized task$_3$’, it is hard to find its physical meaning because the intention of this merged task would be different from the original intention supported by the three successive subtasks ('Proceduralized task$_1$', 'Proceduralized task$_2$', and 'Proceduralized task$_3$'). For this reason, in this study, this second rule is suggested in order to emphasize the constraint such that only successive tasks should be used for creating new hypothetical tasks.

From the abovementioned two rules, a total of 53 hypothetical proceduralized tasks were identified based on the combination of procedural steps given in Table 2. Detailed information about the 53 hypothetical proceduralized tasks is summarized in the *Appendix* of this paper.

## 4. Estimation of empirical human error probability

As stated in Section 1, the purpose of this paper is to empirically estimate HEPs based on TACOM scores. To this end, it is prerequisite to corroborate the fact that there is a pertinent relationship between TACOM scores and the occurrence of human errors. For this reason, human error fractions summarized in the *Appendix* are compared with the associated TACOM scores (Fig. 5).

From Fig. 5, it is expected that the higher the TACOM score increases, the more human operators are apt to make an error including EOO and EOC. This tendency seems to follow an exponential function, which is compatible with a result reported by Ref. [13]. This implies that the estimation of empirical HEPs with respect to TACOM scores is feasible, and hence the empirical HEPs of both EOCs and EOOs are quantified using Logistic regression technique.

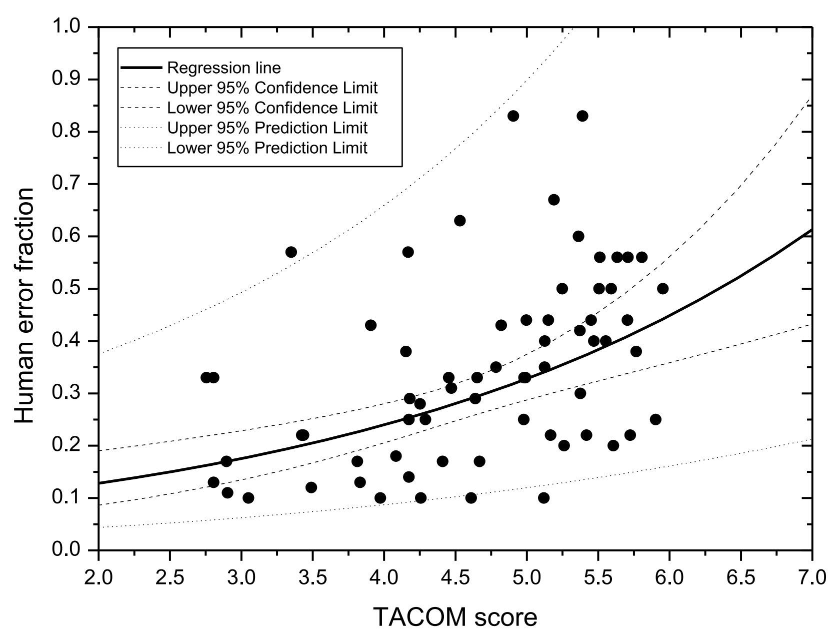
Similar to other regression techniques (e.g., the non-linear regression model illustrated in Fig. 5), the primary purpose of Logistic regression is to predict the correlation between a dependent variable and one or more independent variables. However, one of the unique features expected from the Logistic regression technique is that it can be applied to the analysis of a dependent variable with binary (dichotomous) states, such as *Win/Loss*, *Alive/Dead*, and *Failure/Success* (or *Error/No error*). In addition, since Logistic regression technique considers the concept of *Logit* (the fraction of occurrence probabilities related to binary states), it is expected that the occurrence probability of an EOC (and EOO) can be empirically estimated when the TACOM score plays as the independent variable.

In this light, using R and RStudio [33,41], Logistic regression technique is applied to analyze the correlation between human errors (EOC and EOO) and TACOM scores summarized in the *Appendix* of this paper. Table 3 lists the key parameters of Logistic regression model for both EOC and EOO, and Fig. 6 compares the occurrence probabilities of both EOC and EOO, which are predicted based on the parameter values of *Intercept* and *TACOM* in Table 3.

From Table 3, it seems that both Logistic regression models

**Table 3**
Key parameters for the Logistic regression model of EOC and EOO.

| Parameter | EOC | EOO |
|---|---|---|
| Intercept | −7.228 | −2.359 |
| TACOM | 1.023 | 0.240 |
| Degree of freedom (Total) | 609 | 609 |
| Degree of freedom (Residual) | 608 | 608 |
| Null deviance | 387.7 | 649.8 |
| Residual deviance | 358.3 | 645.3 |




$$y = 0.0686 \cdot e^{0.3129 \cdot x}, R^2 = 0.2365$$

| Item | Degree of freedom | Sum of squares | Mean square | F statistic* |
|---|---|---|---|---|
| Model | 1 | 0.9342 | 0.9342 | 19.8227 |
| Error | 64 | 3.0162 | 0.0471 | |
| Total | 65 | 3.9504 | | |

*$p < 0.01$

**Fig. 5.** Regression analysis between human error frequencies and the associated TACOM scores.

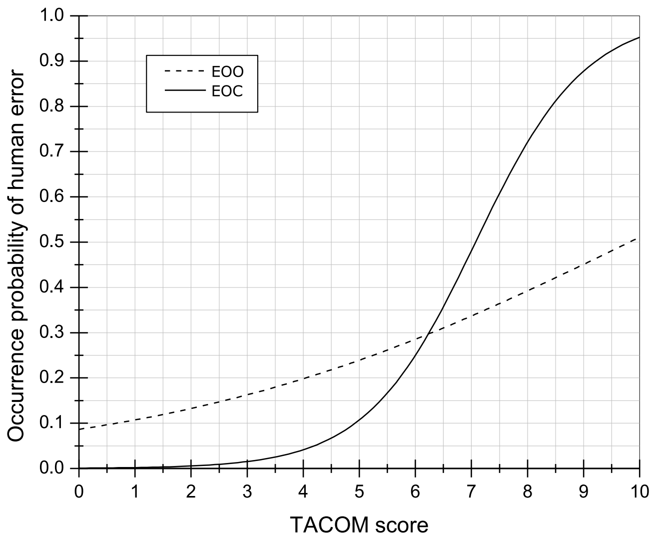



**Fig. 6.** Comparing Logistic regression lines for EOC and EOO.

depicted in Fig. 6 are appropriate [34] because their residual deviances are below 666.5, which corresponds to the critical chi-square value given the probability of 5% and 608 degree of freedom [35]; [40]. In other words, since a high deviance value indicates that a Logistic regression model does not fit the applied data, it can be said that Logistic regression models for both EOC and EOO are statistically relevant for representing their occurrence probabilities based on TACOM scores. Nevertheless, it is also true that Logistic regression model for EOC is better than that for EOO because the residual deviance of the former is almost half that of the latter. Therefore, it is still prudential whether or not TACOM scores are meaningful for explaining the change of occurrence probabilities pertaining to EOO.

In this regard, one promising approach is to see whether or not the residual deviance value decreases after eliminating EOOs of which the main drivers belong to the sub-categories of *Task contents* and *Task type/attribute*. That is, if an EOO occurred because of other reasons that are not related to one of the five sub-measures depicted in Fig. 2, it is hard to directly correlate the occurrence of EOOs with the associated TACOM scores. Therefore, if the residual deviance decreases when these irrelevant EOOs are removed, it is reasonable to expect that TACOM scores can be used for estimating the occurrence probability of EOOs. For this reason, the context information of EOOs observed from the 13 procedural steps in Table 2 was reviewed in detail.

As a result, two interesting contexts were identified from the EOOs that occurred during the performance of the 5th and 6th procedural steps shown in Table 2. The first one is the consideration of a specific constraint given by *Warning*, *Caution* and *Note* statements. According to U.S. Nuclear Regulatory Commission (USNRC), they need to have dedicated features such that:

"Warnings and cautions are derived initially from technical guidelines. They contain information used to prevent actions by control room operators which could injure plant personnel, damage equipment, or endanger public health and safety. [...] They should not contain operator actions. [...] Note statements provide operators with supplemental information concerning specific steps or sequences of steps in the EOP. These statements should [...], and be located so as to ensure that they can easily relate the note to the step or steps to which it applies. Because they are supplemental, notes should not direct operators to perform actions." [36; p. 24].

In short, although *Warning*, *Caution* and *Note* statements contain specific actions that are important for the safety of NPPs, they are not explicitly described as a part of a procedural step. For example, let us assume that the following *Caution* statement is given pertaining to the procedural step shown in Fig. 1: "In order to provide additional water, it is necessary to start Pump_4 if the water level of Tank_1 is less than 30%." With this *Caution* statement, an EOO should be marked if human operators did not start Pump_4 when the water level of Tank_1 was 25%. Indeed, 2 out of 3 EOOs that occurred in the 5th procedural step of Table 2 were marked due to the omission of required actions manifested in a *Caution* statement. Nevertheless, no contents of any *Caution* statements were considered in the quantification of TACOM scores because they are not a part of the procedural step (i.e., do not instruct what has to be done by human operators). This implies that, without the involvement of *Caution* statements, the TACOM scores of certain proceduralized tasks could be underestimated.

The second interesting context recognized from the EOOs observed from the 6th procedural step of Table 2 is related to its execution sequence. In general, it is expected that each procedural step included in a procedure will be carried out in accordance with a predefined execution sequence (e.g., Step 1 → Step 2 → Step 3). However, as briefly explained in Section 2.1, it is difficult to cope with the whole spectrum of dynamically varying situations using a series of prescribed procedural steps. For this reason, in the case of EOPs in NPPs, there are several procedural steps of which the execution sequence has unique features. A non-sequential procedural step is a typical one because it can be carried out at various intervals throughout an EOP [36]. That is, human operators can conduct a non-sequential procedural step at any time when a trigger condition is met.

For example, if the procedural step shown in Fig. 1 corresponds to a non-sequential procedural step, human operators should continuously monitor the water level of Tank_1 so that they can conduct all of the required actions belonging to it regardless of a predefined execution sequence. This implies that human operators are apt to omit required actions if they failed to continuously monitor the key symptoms that are necessary to decide the satisfaction of a trigger condition (e.g., the water level of Tank_1). Here, it should be noted that the characteristic of a non-sequential procedural step belongs to neither *Task contents* nor *Task type/attribute* sub-categories that are preliminarily covered by the TACOM measure. In other words, in terms of generating Logistic regression model, it seems to be inappropriate to involve EOOs caused by the failure of monitoring relevant symptoms. From this concern, since all EOOs observed from the 6th procedural step of Table 2 were caused by this type of failure, it is reasonable to compare TACOM scores with the occurrence probability of EOOs after eliminating those observed from the 6th procedural step.

Actually, the residual deviance of Logistic regression model that was recalculated after eliminating the EOOs pertaining to the abovementioned first and second context from the *Appendix* of this paper is 458.48 with 495 degree of freedom. This newly calculated value is about 30% lower than that of the original value shown in Table 3 (i.e., 645.3). Therefore, it is anticipated that the goodness of fit related to Logistic regression line for EOO will increase if we are further able to pick out EOOs that were directly related to the complexity of proceduralized tasks.

## 5. Discussion and conclusion

As mentioned in Section 1, the reliability of human operators is one of the determinants to ensure the operational safety of socio-technical systems, including NPPs. This implies that understanding when and why the degradation of human performance occurred would be the very first step to enhance the reliability of human operators. In this study, the occurrence probability of human error was empirically investigated with respect to TACOM scores that can represent the complexity levels of proceduralized tasks. To this end, human errors stored in the HuREX database were revisited, which were originally collected from the full-scope training simulator of Korean domestic NPPs equipped with an analog MCR. After that, Logistic regression technique was applied in order to empirically estimate the occurrence probabilities of both EOCs and EOOs with respect to TACOM scores.

As a result, it was shown that Logistic regression models created for estimating these occurrence probabilities are statistically relevant. In particular, compared with EOOs, it seems that the occurrence probability of EOCs is more dependent on TACOM scores. One promising assumption to explain this result is that the occurrence of EOOs largely relies on the dynamic characteristics of an accident condition (i.e., *System* PIF group) rather than the complexity of proceduralized tasks (i.e., *Task* PIF group). In other words, compared with the performance of complicated tasks, human operators are apt to omit detailed actions prescribed in a procedural step when they are faced with an accident condition during which many kinds of critical process parameters are rapidly changing. Indeed, this assumption is conceivable because the natures of the accident scenarios simulated in the full-scope simulator are highly complicated and cognitively challenging.

That is, since the simulation scenario of MSLB followed by SGTR given in Table 1 is the combination of two representative accident conditions, a large number of process parameters concurrently changed with in a very short time period. In addition, due to the rapid progression of this scenario, human operators have to accomplish proceduralized tasks in parallel with the monitoring of the status of key process parameters such as the pressure and temperature of steam generators, which become the trigger condition of non-sequential procedural steps. Similarly, since the ISLOCA scenario was initially designed to push substantial cognitive demand on human operators [37], they have to collect a lot of relevant symptoms in order to properly cope with the on-going accident situation at hand. Accordingly, in terms of EOOs, human operators may not only omit a complicated proceduralized task when they have to conduct it but also fail to decide the necessity for conducting a required action due to rapidly changing situations.

At the same time, if human operators have to use a procedure, it is possible to assume that the effect of such dynamic characteristics on the occurrence of EOCs could be smaller than that of EOOs. This is because an EOC can occur only when human operators first recognized that there is something to be done. If so, it is reasonable to say that the main driver of an EOC is a misinterpretation (or misunderstanding) of a complicated procedural step, which seems to be directly affected by detailed PIFs belonging to both *Task contents* and *Task type/attribute*. Unfortunately, it is not possible to manifest the abovementioned assumption because the occurrence probability of both EOCs and EOOs can be simultaneously affected by the dynamic characteristics. In other words, the occurrence probability estimated in this study (refer to Fig. 6) corresponds to the *basic HEP* containing the effect of PSFs, which has to be distinguished from the *nominal HEP*.

Park and Jung [38] stated that “Nominal HEP is the probability of a given human error when the effects of performance shaping factors (PSFs) have not been considered. In order to yield more realistic human reliability analysis, the nominal HEPs of task elements must be modified according to the task situation” [38, p. 325]. In addition, they also specified that “In general, since industrial conditions can vary in the quality levels of PSFs, a nominal HEP must be modified if PSFs in a specific task situation are not average conditions. Its modified form is the basic HEP, which is the probability of a human error on a task that is considered as an isolated entity, unaffected by any other task” [38, p. 326]. This implies that comparing TACOM scores (i.e., the level of a task complexity) with the occurrence probabilities of both EOCs and EOOs is necessary after eliminating the effect of dynamic characteristics on the occurrence probabilities. For example, when human operators are exposed to such cognitive demanding situations generated by MSLB followed by SGTR or ISLOCA, it is natural to anticipate that the occurrence probability of EOOs will increase due to additional task-specific situations including (1) *Warning*, *Caution* and *Note* statements, and (2) non-sequential procedural steps. This is because the amount of cognitive resources available to human operators is highly limited in such situations, which is crucial for considering these constraints. In addition, it is reasonable to expect that the occurrence probability of EOCs will go up due to a lack of cognitive resources.

Therefore, if we are able to get the occurrence probability of both EOCs and EOOs from a specific situation in which the effect of such dynamic characteristics can be minimized, it is possible to estimate the modified occurrence probabilities of both EOCs and EOOs that seem to be closer to their nominal HEPs. Once the modified occurrence probabilities are available, it is expected that the correlation between the occurrence probabilities and TACOM scores will become more evident.

One promising solution for this issue is to reanalyze human errors observed from the full-scope training simulator of a fully-digitalized MCR equipped with a computerized procedure system (CPS). This is because one of the primitive CPS functions is to support the status monitoring of non-sequential procedural steps [39]. That is, if the CPS automatically checks the trigger condition of non-sequential procedural steps, it is possible to observe its impact on the occurrence of human errors (both EOCs and EOOs). If so, the correlation between TACOM scores and the occurrence probability of human errors would strengthen.

Once we achieve a reliable correlation between TACOM scores and the occurrence probability of human errors, it will also be possible to designate a certain threshold from which the performance of human operators starts to be drastically degraded. Fig. 7 is helpful for explaining this expectation, showing the derivative of Logistic regression line for the EOC in Fig. 6.

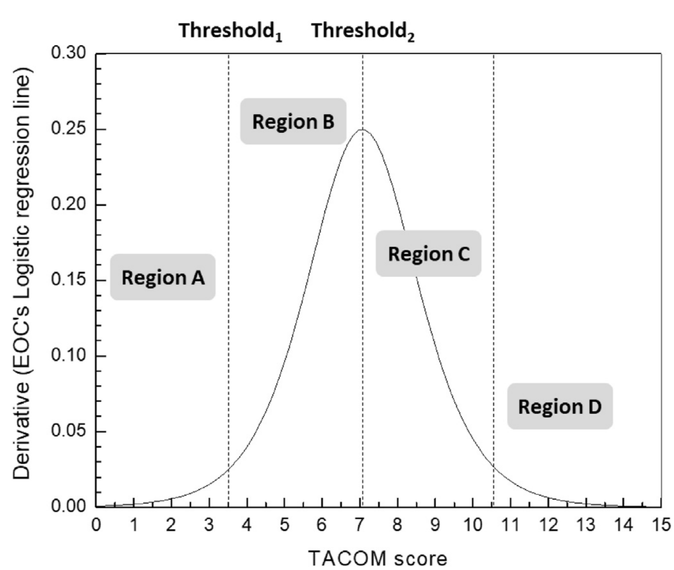



**Fig. 7.** Variation of the occurrence probability of EOC with hypothetical thresholds based on TACOM scores.

As can be seen from Fig. 7, in the case of *Region A* in which TACOM scores are relatively low, the occurrence probability of an EOC appears to slowly increase with an increase in TACOM scores. This tendency seems to radically change, however, when the TACOM score exceeds a specific value ($\text{Threshold}_1$) that distinguishes *Region A* and *Region B*. This implies that the complexity of a proceduralized task significantly affects the occurrence probability of an EOC. Interestingly, this tendency quickly disappears and goes down in *Region C* after the TACOM score passes another particular value ($\text{Threshold}_2$). Finally, in *Region D*, the occurrence probability of an EOC seems to slowly decrease according to an increase in TACOM score.

Although there are many technical issues to be resolved before suggesting the abovementioned thresholds, it is strongly anticipated that the TACOM measure could play a crucial role in accomplishing this goal. The result of this study is meaningful because it would become the starting point of this research direction.

## 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.

## Acknowledgment

This work was supported by a grant from the Nuclear Research & Development Program of the National Research Foundation of Korea, funded by the Korean government, Ministry of Science, ICT & Future Planning (grant number 2019M2C9A1055906), and by a Nuclear Safety Research Program grant from the Korea Foundation of Nuclear Safety (KOFONS), funded by the Nuclear Safety and Security Commission of the Republic of Korea (No. 2105029).

## Appendix. Extension of Table 2 with hypothetical proceduralized tasks

| ID* | Trial | Number of EOOs | Number of EOCs | TACOM score | Human error fraction |
|---|---|---|---|---|---|
| 1 | 17 | 1 | 2 | 4.084 | 0.18 |
| 2 | 18 | 3 | 0 | 3.813 | 0.17 |
| 3 | 18 | 2 | 0 | 2.904 | 0.11 |
| 4 | 10 | 1 | 0 | 3.050 | 0.10 |
| 5 | 9 | 3 | 0 | 2.806 | 0.33 |
| 6 | 9 | 3 | 0 | 2.755 | 0.33 |
| 7 | 9 | 1 | 1 | 3.424 | 0.22 |
| 8 | 9 | 2 | 0 | 3.436 | 0.22 |
| 9 | 9 | 2 | 2 | 5.151 | 0.44 |
| 10 | 8 | 1 | 0 | 2.806 | 0.13 |
| 11 | 8 | 0 | 1 | 3.832 | 0.13 |
| 12 | 7 | 0 | 4 | 3.350 | 0.57 |
| 13 | 6 | 1 | 0 | 2.896 | 0.17 |
| H1 | 12 | 1 | 2 | 4.979 | 0.25 |
| H2 | 17 | 4 | 2 | 4.784 | 0.35 |
| H3 | 18 | 5 | 0 | 4.252 | 0.28 |
| H4 | 17 | 2 | 0 | 3.490 | 0.12 |
| H5 | 10 | 1 | 0 | 3.974 | 0.10 |
| H6 | 8 | 3 | 0 | 4.153 | 0.38 |
| H7 | 8 | 2 | 0 | 4.174 | 0.25 |
| H8 | 7 | 3 | 0 | 3.907 | 0.43 |
| H9 | 7 | 0 | 1 | 4.174 | 0.14 |
| H10 | 7 | 2 | 0 | 4.180 | 0.29 |
| H11 | 9 | 2 | 2 | 5.450 | 0.44 |
| H12 | 8 | 1 | 1 | 4.290 | 0.25 |
| H13 | 7 | 0 | 4 | 4.169 | 0.57 |
| H14 | 6 | 1 | 0 | 4.410 | 0.17 |
| H15 | 9 | 0 | 2 | 5.167 | 0.22 |
| H16 | 12 | 3 | 2 | 5.372 | 0.42 |
| H17 | 16 | 5 | 2 | 4.997 | 0.44 |
| H18 | 16 | 5 | 0 | 4.471 | 0.31 |
| H19 | 10 | 1 | 0 | 4.258 | 0.10 |
| H20 | 8 | 5 | 0 | 4.532 | 0.63 |
| H21 | 6 | 2 | 0 | 4.652 | 0.33 |
| H22 | 6 | 2 | 0 | 4.453 | 0.33 |
| H23 | 6 | 0 | 1 | 4.669 | 0.17 |
| H24 | 7 | 2 | 0 | 4.639 | 0.29 |
| H25 | 5 | 1 | 0 | 5.606 | 0.20 |
| H26 | 9 | 0 | 2 | 5.418 | 0.22 |
| H27 | 9 | 3 | 2 | 5.512 | 0.56 |
| H28 | 12 | 4 | 2 | 5.506 | 0.50 |
| H29 | 17 | 4 | 2 | 5.125 | 0.35 |
| H30 | 10 | 1 | 0 | 4.610 | 0.10 |
| H31 | 6 | 5 | 0 | 4.906 | 0.83 |
| H32 | 6 | 2 | 0 | 4.991 | 0.33 |
| H33 | 7 | 2 | 1 | 4.821 | 0.43 |
| H34 | 6 | 2 | 0 | 4.981 | 0.33 |

(*continued*)

| ID* | Trial | Number of EOOs | Number of EOCs | TACOM score | Human error fraction |
|---|---|---|---|---|---|
| H35 | 9 | 0 | 2 | 5.724 | 0.22 |
| H36 | 9 | 2 | 2 | 5.705 | 0.44 |
| H37 | 9 | 3 | 2 | 5.633 | 0.56 |
| H38 | 12 | 4 | 2 | 5.591 | 0.50 |
| H39 | 10 | 1 | 2 | 5.375 | 0.30 |
| H40 | 10 | 1 | 0 | 5.119 | 0.10 |
| H41 | 6 | 4 | 0 | 5.190 | 0.67 |
| H42 | 6 | 2 | 1 | 5.248 | 0.50 |
| H43 | 5 | 2 | 0 | 5.126 | 0.40 |
| H44 | 5 | 1 | 0 | 5.261 | 0.20 |
| H45 | 8 | 0 | 2 | 5.903 | 0.25 |
| H46 | 8 | 2 | 2 | 5.953 | 0.50 |
| H47 | 9 | 3 | 2 | 5.806 | 0.56 |
| H48 | 9 | 3 | 2 | 5.708 | 0.56 |
| H49 | 8 | 1 | 2 | 5.766 | 0.38 |
| H50 | 10 | 2 | 2 | 5.554 | 0.40 |
| H51 | 6 | 4 | 1 | 5.390 | 0.83 |
| H52 | 5 | 2 | 0 | 5.470 | 0.40 |
| H53 | 5 | 3 | 0 | 5.362 | 0.60 |

*H denotes *Hypothetical proceduralized task* expended along with the arbitrary combination of procedural steps that successively occur in a row.

## References


[1] G. Baxter, I. Sommerville, Socio-technical systems: from design methods to systems engineering, Interact. Comput. 23 (1) (2011) 4–17.
[2] F.E. Emery, E.L. Trist, Socio-technical systems, management sciences: models and techniques, in: C.W. Churchman, M. Verhulst (Eds.), Proceedings of the Sixth International Meeting of the Institute of Management Sciences, Paris, 7-11 September 1959, vol. 2, 1960. France.
[3] Oxford Reference, Available at: https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100515814, 2021.
[4] C.D. Griffith, S. Mahadevan, Inclusion of fatigue effects in human reliability analysis, Reliab. Eng. Syst. Saf. 96 (11) (2011) 1437–1447.
[5] International Atomic Energy Agency (IAEA), Operating experience with nuclear power stations in member countries (2020 edition), Available at: https://www.iaea.org/publications/14782/operating-experience-with-nuclear-power-stations-in-member-states, 2020.
[6] A. Petrillo, F. Zomparelli, The importance of human error and reliability management in critical conditions and infrastructures, in: F. De Felice, A. Petrillo (Eds.), Human Factors and Reliability Engineering for Safety and Security in Critical Infrastructures: Decision Making, Theory, and Practice, Springer-Verlag, 2018.
[7] Nuclear Energy Agency (NEA), Nuclear Power Plant Operating Experiences from the IAEA/NEA Incident Reporting System 2015-2017, NEA No. 7482, 2020, p. 2020.
[8] J. Park, W. Jung, J. Kim, Inter-relationships between performance shaping factors for human reliability analysis of nuclear power plants, Nucl. Eng. Technol. 52 (1) (2020) 87–100.
[9] J. Park, Scrutinizing inter-relations between performance influencing factors and the performance of human operators pertaining to the emergency tasks of nuclear power plants – an explanatory study, Ann. Nucl. Energy 38 (2011) 2521–2532.
[10] J. Park, The Complexity of Proceduralized Tasks, Springer-Verlag, Berlin, Germany, 2009.
[11] L. Podofillini, J. Park, V.N. Dang, Measuring the influence of task complexity on human error probability: an empirical evaluation, Nucl. Eng. Technol. 45 (2) (2013) 151–164.
[12] J. Park, Investigating the TACOM measure as a general tool for quantifying the complexity of procedure guided tasks, Reliab. Eng. Syst. Saf. 129 (2014) 66–75.
[13] I. Jang, Y. Kim, J. Park, Investigating the effect of task complexity on the occurrence of human errors observed in a nuclear power plant full-scope simulator, Reliab. Eng. Syst. Saf. 214 (2021) 107704.
[14] Health and Safety Executive, Human factors: managing human failures, Available at: https://www.hse.gov.uk/humanfactors/topics/humanfail.htm, 2021.
[15] P. Kumar, S. Gupta, Y.R. Gunda, Estimation of human error rate in underground coal mines through retrospective analysis of mining accident reports and some error reduction strategies, Saf. Sci. 123 (2020) 104555.
[16] S. Wenwen, J. Fuchuan, Z. Qiang, C. Jingjing, Analysis and control of human error, Procedia Eng. 26 (2011) 2126–2132.
[17] J.W. Hendricks, S.C. Peres, Beyond human error: an empirical study of the safety Model 1 and Model 2 approaches for predicting workers' behaviors and outcomes with procedures, Saf. Sci. 134 (2021) 105016.
[18] S. Massaiu, Human Error Models and Behaviors, IFE/HR/E-2005/031, Institute for Energy Technology, Norway, 2005.
[19] P. Li, Y. Wang, J. Chen, Z. Luo, L. Dai, An experimental study on the effects of task complexity and knowledge and experience level on SA, TSA and workload, Nucl. Eng. Des. 376 (2021) 111112.
[20] P. Liu, Z. Li, Comparison of task complexity measures for emergency operating procedures: convergent validity and predictive validity, Reliab. Eng. Syst. Saf. 121 (2014) 289–293.
[21] X. Zhao, R. Zhang, S. Zhao, Y. Wu, Z. He, Incorporation TACOM and SPAR-H into the operating procedure of nuclear power plants, Ann. Nucl. Energy 114 (2018) 451–457.
[22] Y.H.J. Chang, Y. Kim, J. Park, L. Criscione, SACADA and HuREX Part 1: the Use of SACADA and HuREX Systems to Collect Human Reliability Data. Under Review in Nuclear Engineering and Technology, 2021 (as of April 2021).
[23] W. Jung, J. Park, Y. Kim, S.Y. Choi, S. Kim, HuREX – a framework of HRA data collection from simulators in nuclear power plants, Reliab. Eng. Syst. Saf. 194 (2020) 106235.
[24] Y. Kim, Considerations for generating meaningful HRA data: lessons learned from HuREX data collection, Nucl. Eng. Technol. 52 (2020) 1697–1705.
[25] J. Park, Y. Kim, W. Jung, A Framework to Estimate HEPs from the Full-Scope Simulators of NPPs: Unsafe Act Definition, Identification and Quantification, KAERI/TR-6401, 2016. Daejeon, Republic of Korea.
[26] Korea Hydro and Nuclear Power company, Evaluation of Human Error Probabilities in Digital Environment and the Reliability of Safety Critical Software, 2019-50003339-0000TR (A16IP42), Daejeon, Republic of Korea, 2019.
[27] R.L. Boring, K.D. Thomas, T.A. Ulrich, R.T. Lew, Computerized operator support systems to aid decision making in nuclear power plants, Procedia Manuf. 3 (2015) 5261–5268.
[28] S. Park, J. Park, G. Heo, Transient diagnosis and prognosis for secondary system in nuclear power plants, Nucl. Eng. Technol. 48 (2016b) 1184–1191.
[29] M. Hsieh, S. Hwang, K. Liu, S.M. Liang, C. Chuang, A decision support system for identifying abnormal operating procedures in a nuclear power plant, Nucl. Eng. Des. 249 (2012) 413–418.
[30] J. Kim, W. Jung, Y.S. Son, The MDTA-based method for assessing diagnosis failures and their risk impacts in nuclear power plants, Reliab. Eng. Syst. Saf. 93 (2008) 337–349.
[31] B. Reer, V.N. Dang, S. Hirschberg, The CESA method and its application in a plant-specific pilot study on errors of commission, Reliab. Eng. Syst. Saf. 83 (2004) 187–205.
[32] J. Park, I. Jang, Y. Kim, Feasibility of the TACOM measure as a yardstick to identify the level of task complexities under a procedure guided task environment, in: 2021 International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA 2021), November 7–12, 2021 (Virtual Meeting).
[33] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2020. Available at: https://www.R-project.org/.
[34] J.M. Hilbe, Practical Guide to Logistic Regression, CRC Press, 2015.
[35] Omnicalculator, Available at: https://www.omnicalculator.com/statistics/critical-value, 2021.
[36] US Nuclear Regulatory Commission, Guidelines for the Preparation of Emergency Operating Procedures: Resolution of Comments on NUREG-0799, NUREG-0899, Washington, D. C, 1982.
[37] E.M. Roth, R.J. Mumaw, P.M. Lewis, An Empirical Investigation of Operator Performance in Cognitively Demanding Simulated Emergencies, NUREG/CR-6208, U.S. Nuclear Regulatory Commission, Washington, D. C, 1994.
[38] K.S. Park, K.T. Jung, Considering performance shaping factors in situation-specific human error probabilities, Int. J. Ind. Ergon. 18 (4) (1996) 325–331.
[39] J. Hong, M. Lee, D. Hwang, Computerized procedure system for the APR1400 simulator, Nucl. Eng. Des. 239 (12) (2009) 3092–3104.
[40] Medical, Available at: https://www.medcalc.org/manual/chi-square-table.php, 2021.
[41] RStudio Team, RStudio, Integrated development for R. RStudio, Inc., Boston, MA, 2019. Available at: http://www.rstudio.com/.