International Journal of Industrial Ergonomics 39 (2009) 756–765

Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

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

A spaceflight operation complexity measure and its experimental validation

Yijing Zhang [a,b,*], Zhizhong Li [a], Bin Wu [b], Su Wu [a]

[a] Department of Industrial Engineering, Tsinghua University, Beijing 100084, PR China
[b] Laboratory of Astronaut Selection and Training, China Astronaut Research and Training Center, Beijing 100094, PR China

ARTICLE INFO

Article history:
Received 19 March 2008
Received in revised form
3 February 2009
Accepted 14 March 2009
Available online 17 April 2009

Keywords:
Operation complexity
Operation performance
Spaceflight

ABSTRACT

A measure of operation complexity in spaceflight is proposed using a weighted Euclidean norm based on four factors: complexity of operation step size (COSS), complexity of operation logic structure (COLS), complexity of operation instrument information (COII), and complexity of space mission information (CSMI). The development of the operation complexity measure followed four steps. First, four factors were identified to be reflected in the operation complexity measure for spaceflight. Second, the entropy theory was adopted to measure the four factors. Then, the weights of the four factors were determined based on a questionnaire survey of 10 astronauts. Finally, the operation complexity values of spaceflight operations were determined by the weighted Euclidean norm of the four factors. To verify the validity of this complexity measure, a one-factor experiment was designed to test the proposed hypotheses. Ten subjects participated in the experiment and performed 179 trials. Both objective indexes (operation time and error rate) and subjective indexes (workload evaluated by NASA Task Load Index questionnaire and subjective complexity rating) were used in the experiment. The data analysis showed that the average operation time, subjective complexity rating, and subjective workload could be predicted well from the operation complexity value ($R = 0.876$, 0.802, and 0.698, respectively); and the error rate could only be partly explained by the operation complexity value ($R = 0.343$).

Relevance to industry: The proposed operation complexity measure can be used for ergonomics evaluation of spaceflight operation design. It can also be used for astronaut training planning. Training resources can be allocated to spaceflight tasks according to their operation complexity.

1. Introduction

The importance of spaceflight and exploration programs and the activities of astronauts in space have been increasingly recognized worldwide. Effective astronaut operation can improve the reliability and safety of manned spaceflight systems. On the other hand, human error will reduce the reliability of a system, and even lead to mission failure or catastrophic accidents (Wang and Huang, 1996; Nelson et al., 1998; Nelson, 1999). From 1971 to 1997, there have been eight accidents involving spacecraft rendezvous and docking (RVD), a critical spaceflight operation. Among the eight accidents, five were due to human errors (Shayler, 2005). In order to ensure the safety and reliability of manned spaceflight, much research on countermeasures have been carried out (Ostrom et al., 1992; Pierre and Gregory, 1995; Seastrom et al., 2004). Attention has been paid to the design of aerospace systems and equipment (Pierre and Gregory, 1995; Seastrom et al., 2004). Only a few papers, such as the report by Ostrom et al. (1992), directly addressed human reliability. From the analysis of RVD mission failures, it has been found that astronaut training is critical to improve the reliability of manned spaceflight systems (Shayler, 2005).

In astronaut training, it is unreasonable and costly to assign the same time and resources to different operations. It would be more reasonable to allocate time for an operation according to its complexity.

There have already been studies on complexity measures. Xing and Manning (2005) reviewed the literature on complexity including articles on general concepts, information complexity (Xing, 2004), cognitive complexity (Rauterberg, 1992), and display complexity. Davis and LeBlanc (1988) proposed the notion of computer program complexity and compared several complexity measures. While these studies were focused on different areas, they all agreed on three factors associated with complexity: the quantity of basic information elements (size), the variety of elements (variety), and the relationship between elements (rule) (Xing and Manning, 2005).

In software engineering, several complexity measures have been studied and developed. The entropy concept has been found

---

[*] Corresponding author. Laboratory of Astronaut Selection and Training, China Astronaut Research and Training Center, Beijing 100094, PR China. Tel./fax: +86 10 68746545.

E-mail address: zyj.acc@gmail.com (Y. Zhang).

0169-8141/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ergon.2009.03.003

to be able to act as a unified software complexity measure taking into account different dimensions and factors (Gonzalez, 1995). Actually, the entropy concept has commonly been used and tested in various complexity measures (Davis and LeBlanc, 1988; Lew et al., 1988; Park et al., 2001a). The above three aspects of complexity were used to analyze the variety of chunks in a program (Davis and LeBlanc, 1988), the complexity characteristics of data structures (Lew et al., 1988), and the step complexity of emergency operating procedures in nuclear power plants (Park et al., 2001a).

Though they can be analyzed from the similar complexity aspects, the operations in spaceflight are quite different from those in nuclear plants. The former is performed in a small cabin under a situation of microgravity and isolation. Under those conditions, the layout and method of instrument manipulation are important factors for the operation complexity. In addition, under the guidance of grand control support teams, astronauts have limited independent decision-making opportunities, but they do need to know many things about the mission, such as the function and method of operation for every subsystem.

Thus, a way to measure operation complexity in spaceflight was preliminarily proposed in a symposium (Zhang et al., 2007), to reflect the special situation in spaceflight operation. This paper extensively describes the operation complexity measure in spaceflight and is also aimed at examining how operation complexity correlates with the performance of operators. A one-factor experiment was designed to test the proposed hypotheses.

The remainder of this paper is organized as follows. Section 2 describes the measure for operation complexity in spaceflight. The experiment design is described in Section 3. The experiment results and data analysis are presented in Section 4. The last section concludes this paper.

2. Method

2.1. Spaceflight operation complexity measurement

One spaceflight mission is composed of many operation units, one of which is an action group to achieve a defined purpose by using human–machine interface, such as cabin environment control. A spaceflight operation complexity measure was developed to evaluate the design of the human–machine interface and spaceflight operations. Thus, the complexity measure should reflect the characteristics of the interface and operations. On the other hand, if the interface and operations are well designed, we can expect better operation performance; vice versa. In other words, the complexity measure should have a good correlation with operation performance. Therefore, the complexity measure could be validated by correlation/regression analysis between complexity values and performance values. The development of the operation complexity measure followed four steps. First, four factors were identified to be reflected in the operation complexity measure for spaceflight. Second, the entropy theory was adopted to measure the four factors. Then, the weights of the four factors were determined based on a questionnaire survey of astronauts. Finally, the operation complexity values of spaceflight operations were determined by the weighted Euclidean norm of the four factors.

2.1.1. Identification of operation complexity factors

Edmonds (1999) indicated that the complexity of things depends on which aspect you are concerned with. In astronaut training on land, one of the main goals is to evaluate the design of the human–machine interface and the spaceflight operations, which is also the objective of the complexity study. The factors influencing spaceflight operations were identified by a survey among 10 astronauts and 10 trainers of astronauts. All of the 20 subjects thought that the length of operations, the logic structure of operations, the task information for an operation, and the type and number of monitors and controllers would affect the precision and reliability of operations in spaceflight. Moreover, eight of the 10 astronauts considered that an operation may also be influenced by other factors, such as microgravity, small space, and personality. The influences of these factors on individual spaceflight operations are complicated and are not about the interfaces and operations to be evaluated, so they were not considered in operation complexity measurement.

Finally, the following four factors were selected to describe the operation complexity: the complexity of operation step size (COSS), which evaluates the amount of actions contained in one operation unit, the complexity of operation logic structure (COLS), which describes the logical sequence to conduct the activities of one operation unit, the complexity of operation instrument information (COII), which denotes the type and number of monitors and controllers in one operation unit, and the complexity of space mission information (CSMI), which is related to the difficulty level of the task information for completing one operation unit. In the step complexity proposed by Park et al. (2001a) for nuclear plant emergency procedures, three factors were considered: the amount of operators' actions in each step, the logic structure of each step, and the amount of information in each step. The first two factors considered in spaceflight operation complexity are similar to Park et al.'s step complexity. However, the step complexity for nuclear plants deals with the complexity of written procedures without considering the issues of control interface, i.e. the layout and type of instruments. It was found in ground spaceflight training that the layout and methods of manipulating the interfaces have a demonstrable effect on astronaut performance (Wang et al., 2006). Moreover, under extreme conditions such as the small cabins, microgravity and isolation in spaceflight, interface design becomes more critical to the success of spaceflight missions (Morphew, 2001). So the layout and manipulation methods should be regarded as an indispensable complexity factor in spaceflight operation.

2.1.2. Measurement of operation complexity factors

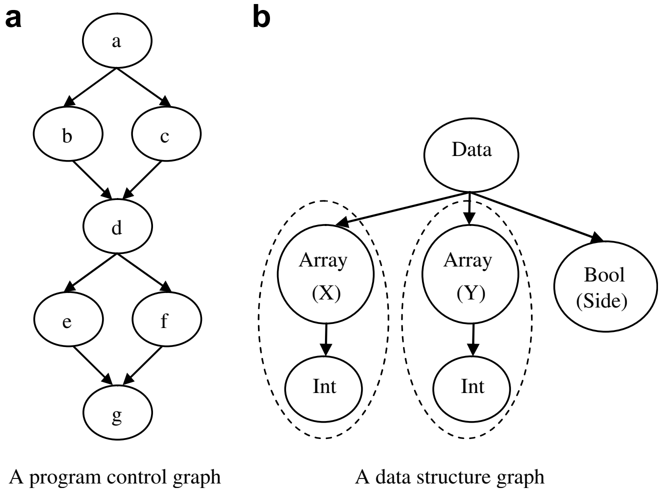
With the advantage of taking into account different complexity dimensions and factors (Gonzalez, 1995), entropy measures were used to quantify the four factors. There are two kinds of entropy measures for a graph in the entropy theory which were described in Mowshowitz's (1967) work: the chromatic information content (or the first-order entropy) and the structural information content (or the second-order entropy). In the case of a computer program control graph, the first-order entropy can be used to evaluate the regularity of the control logic of a given program, while the second-order entropy can be used to evaluate the number of hierarchical levels (or size) of the control graph, and thus can represent the amount of information required to understand the graph.

To calculate the entropy values of a graph, first-order entropy regards the nodes that have the same numbers of incoming arcs and outgoing arcs as equivalent classes. Second-order entropy considers nodes equivalent if they have the same number and type of neighbors (one arc distance). The relative frequency of occurrence of a node type is the number of nodes in the class divided by the total number of nodes in the graph (Davis and LeBlanc, 1988).

The entropy formula used in this paper is shown in Eq. (1), where N is the number of node classes and $P(A_i)$ is the relative frequency of the class A_i:

$$H = -\sum_{i=1}^{N} P(A_i)\log_2 P(A_i) \quad (1)$$

Fig. 1a shows a program control graph that describes the logic sequence of a program (Davis and LeBlanc, 1988). A data structure graph is shown in Fig. 1b, which denotes a data record with the



A program control graph A data structure graph

Fig. 1. An example for entropy calculation.

following structure: Struct Data {int X [10]; int Y [10]; Bool Side} (Lew et al., 1988).

In Fig. 1a, there are four node classes of first-order entropy: {a}, {b,c,e,f}, {d} and {g}. Thus, $N=4$. The probabilities of the four node classes are 1/7, 4/7, 1/7, and 1/7, respectively. They are denoted by $P(A_i)$ in Eq. (1). Thus, the value of first-order entropy of Fig. 1a is:

$$H^1_{\text{Fig.1}} = -\left(\frac{1}{7}\log_2\frac{1}{7}+\frac{4}{7}\log_2\frac{4}{7}+\frac{1}{7}\log_2\frac{1}{7}+\frac{1}{7}\log_2\frac{1}{7}\right) = 1.664.$$

Similarly, there are three node classes in Fig. 1b: {Data}, {Array(X), Array(Y)}, and {int, int, Bool} and the value of first-order entropy is 1.459.

According to the notion of second-order entropy, there are five node classes in Fig. 1a: {a}, {b,c}, {d}, {e,f}, and {g}. The probabilities of the five node classes are 1/7, 2/7, 1/7, 2/7 and 1/7, respectively. So the value of second-order entropy of Fig. 1a is:

$$H^2_{\text{Fig.1}} = -\left(\frac{1}{7}\log_2\frac{1}{7}+\frac{2}{7}\log_2\frac{2}{7}+\frac{1}{7}\log_2\frac{1}{7}+\frac{2}{7}\log_2\frac{2}{7}+\frac{1}{7}\log_2\frac{1}{7}\right) = 2.236$$

Similarly, there are five node classes in Fig. 1b: {Data}, {Array(X), Array(Y)}, {int}, {int}, and {Bool} and the value of second-order entropy is 2.252.

To quantify the four factors for spaceflight operations, three kinds of graphs need to be developed by an operation analysis. They are the operation control graph, the operation instrument information graph, and the mission information hierarchy graph. The operation control graph is in the form of a program control graph and depicts the logic structure to perform a task, while the operation instrument information graph and mission information hierarchy graph are expressed in the form of data structure graphs and demonstrate the instrument information and mission information required to perform a task, respectively. These graphs are described in more detail with examples in Section 2.1.4. The corresponding measures for the four factors are listed in Table 1.

2.1.3. Weighting operation complexity factors

It is not sufficient to simply add up the values of these factors, since the four factors describe the operation complexity from the aspects with different importance. Thus it is more proper to give different weights for the four factors. The determination of the weights of the four factors is an important part of the operation complexity measure. Unfortunately, these weights cannot be calculated objectively and accurately based on scientific principles. Instead, they often have to be determined based on subjective evaluation which is often obtained with some uncertainty. Therefore, the Fuzzy Analytic Hierarchy Process (AHP) method and set-valued statistics were used in this paper to calculate the weights (Zhang et al., 2008).

AHP was first developed by Saaty (1980) and is widely used for multi-criteria decision-making (Saaty, 1988). Traditional AHP cannot express the uncertainty of human judgment. Thus, Fuzzy AHP was developed, in which the uncertainties in comparison are expressed as fuzzy sets or fuzzy numbers, such as “between three and five times less important” (Tang and Beynon, 2005).

Fuzzy AHP was adopted in this study for the determination of weights for the four complexity factors. It follows three steps.

2.1.3.1. Obtaining judgment matrices by pairwise comparisons. Ten Chinese astronauts made pairwise comparisons among the four factors. The Saaty’s 1–9 scales (indicating equal importance to the extreme relative importance of one factor over another factor) were used in these comparisons. To indicate the uncertainty in comparison, three fuzzy levels were set for each pairwise comparison: high certainty (±0.5); medium certainty (±1); and low certainty (±1.5) for the judgment. Through a two-round survey with 10 astronauts, 10 matrices of pairwise comparisons were obtained. An element of the comparison matrix, a_{ij} (i.e. a comparison of the ith factor with the jth factor), is a triangular fuzzy number defined as $a_{ij}=(l_{ij}, m_{ij}, u_{ij})$, where m_{ij}, u_{ij}, and l_{ij} are the median, upper bound, and lower bound values for a_{ij}, respectively (Mikhailov and Tsvetinov, 2004). For example, one expert thinks that the first factor is extremely important (9) than the second factor with high certainty (±0.5), then $a_{12}=(8.5, 9, 9.5)$.

2.1.3.2. Determination of the weights in a single astronaut’s opinion. For two triangular fuzzy numbers: $a_1=(l_1, m_1, u_1)$ and $a_2=(l_2, m_2, u_2)$, the possibility of $(a_1 \geq a_2)$ is defined as V in Eq. (2) (Ji and Zhang, 2005):

$$V(a_1 \geq a_2) = \begin{cases} 1 & m_1 > m_2 \\ \dfrac{l_2-u_1}{(m_1-u_1)-(m_2-l_2)} & m_1 < m_2, l_2 \leq u_1 \\ 0 & \text{others} \end{cases}. \quad (2)$$

According to the comparison matrix of one astronaut, the importance level of the ith factor is defined as S_i in the following Eq. (3):

$$S_i = \sum_{j=1}^{4} a_{ij} \otimes \left(\sum_{i=1}^{4}\sum_{j=1}^{4} a_{ij}\right)^{-1}, \quad (3)$$

so the possibility of the importance of one factor over another factor is shown as d_i in Eq. (4):

$$d_i = \min_{j=1,\ldots,4 \quad j\neq i} V(S_i > S_j) \quad (i = 1,2,3,4). \quad (4)$$

Then, the weight of each factor judged by an astronaut can be determined from Eq. (5).

$$w = (w_1, w_2, w_3, w_4) = \left(\frac{d_1}{\sum_{i=1}^4 d_i}, \frac{d_2}{\sum_{i=1}^4 d_i}, \frac{d_3}{\sum_{i=1}^4 d_i}, \frac{d_4}{\sum_{i=1}^4 d_i}\right). \quad (5)$$

2.1.3.3. Determination of the final weights by set-valued statistics. By the above two steps of Fuzzy AHP, 10 sets of the

Table 1
Operation complexity factors and measures.

Complexity factors	Information graphs	Measures
Complexity of operation step size (COSS)	Operation control graph	The second-order entropy of operation control graph
Complexity of operation logic structure (COLS)	Operation control graph	The first-order entropy of operation control graph
Complexity of operation instrument information (COII)	Operation instrument information graph	The second-order entropy of operation instrument information graph
Complexity of space mission information (CSMI)	Mission information hierarchy graph	The second-order entropy of mission information hierarchy graph

weights judged by 10 astronauts were validated from the 10 comparison matrices $w^{(k)}$ ($k = 1,\ldots,10$). Then, to integrate 10 astronauts' opinions, the 10 sets of weights were computed by the set-valued statistics method, named Hadamard statistics to get the final weights (Wei and Zhou, 1998) as

$$\tilde{w}_i = \left(\prod_{k=1}^{10} \tilde{w}_i^{(k)} \right)^{\frac{1}{10}}, \quad i = 1, \ldots, 4. \tag{6}$$

Then the weights were normalized as

$$w_i = \frac{\tilde{w}_i}{\sum_{i=1}^{4} \tilde{w}_i}, \quad i = 1, \ldots, 4. \tag{7}$$

The final results are given in Fig. 2. The weights for the four factors are 0.1725, 0.3821, 0.1965, and 0.2487, respectively.

The operation complexity can then be determined by the Euclidean norm of the four factors:

$$\mathrm{OC} = \sqrt{(0.1725 \times \mathrm{COSS})^2 + (0.3821 \times \mathrm{COLS})^2 + (0.1965 \times \mathrm{COII})^2 + (0.2487 \times \mathrm{CSMI})^2} \tag{8}$$

2.1.4. Evaluation of operation complexity

The operation complexity evaluation procedure includes the following steps:

(1) Choose one operation unit;
(2) Analyze, collect and summarize the operation information of the unit;
(3) Develop three types of information graphs for the operation unit: operation control graph, operation instrument information graph, and mission information hierarchy graph;
(4) Compute the values of the four factors according to Table 1;
(5) Determine the operation complexity value of the operation unit according to Eq. (8).

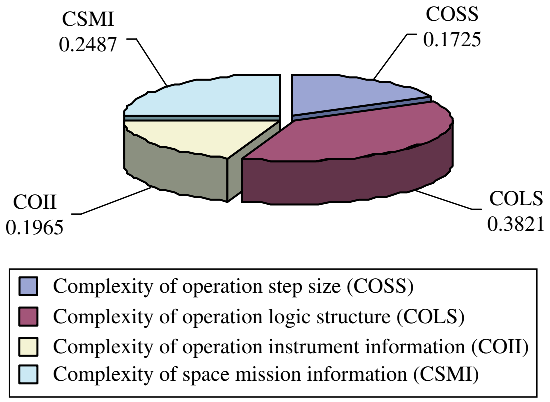


Fig. 2. Weights of four operation complexity factors.

An example of a spaceflight operation unit is discussed here to demonstrate how to evaluate operation complexity. Fig. 3 shows the example, i.e. the operations when the oxygen pressure of the cabin is below the standard level.

For the example, the required mission information, instrument information, and operator actions were obtained from the operation description, spacecraft control interface, and interviews with astronauts and spaceflight experts. This unit was decomposed into 10 individual actions for the operation step size analysis. The mission information analysis involved seven function steps (as shown in Fig. 3). The results of the operation information analysis are listed in Table 2.

Based on Table 2, the operation control graph was developed as a flow chart of actions flow depicting the logic structure of the 10 actions, as shown in Fig. 4.

As shown in Table 2, the operation unit involves five instruments. The O_2 pressure gauge and the whole cabin pressure gauge are located on the first page of the monitoring pages. The process variable is the pressure of oxygen (P). The O_2 flux gauge is located on the second page of the monitoring pages. The process variable is flux (L). Valve 1 has two switches and the process variables of each switch can be considered as Boolean variables with "on/off" values. Valve 2 also has two switches and the process variables are also Boolean variables. The more instruments there are involved in an operation unit, the more complex the operation is. When one instrument is used, it is necessary to know where the instrument is located, how to

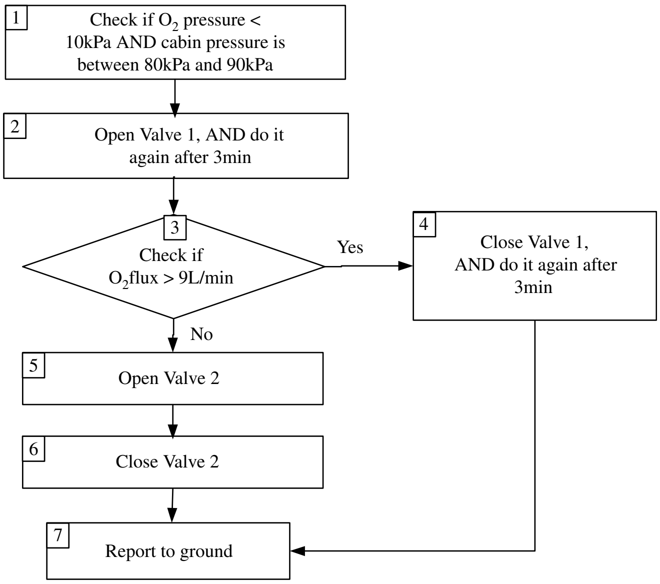


Fig. 3. An operation unit example.

Table 2
Operation information of an operation unit example.

Mission information	Operator actions	Instrument information
1. Know two gauges' state and judge variable values.	1.1. Check: O_2 pressure < 10 kPa 1.2. Check: The cabin pressure is between 80 kPa and 90 kPa	1. O_2 pressure gauge: Location – p1 Variable type – *P* (Pressure)
2. Know valve 1 opening method and remember to do it again.	2.1. Open Valve 1 2.2. Open Valve 1 again after 3 min	2. Cabin pressure gauge: Location – p1 Variable type – *P*
3. Know one gauges' state and judge variable values	3. Check: O_2 flux > 9 L/min	3. O_2 flux gauge: Location – p2 Variable type – *L* (flux)
4. Know Valve 1 closing method and remember to do it again.	4.1. Close Valve 1 4.2. Close Valve 1 again after 3 min	4. Valve 1: Rule – two switches Variable types – *B* (Boolean)
5. Know Valve 2 opening method.	5. Open Valve 2	5. Valve 2: Rule – two switches Variable types – *B*
6. Know Valve 2 closing method.	6. Close Valve 2	
7. Know how to report.	7. Report to ground command	

manipulate the instrument and how to judge the values of the process variables. The operation instrument information graph in a data structure form with the above-mentioned items is illustrated in Fig. 5. Here the S node simply represents the root node; the nodes on the second level represent the instruments of the operation unit, and the downmost nodes represent the datatypes of the corresponding instruments. The location or the manipulated components of an instrument are shown as intermediate nodes.

The mission information represents the amount of information required for each step. According to spacecraft engineering knowledge and previous space operation practice, the mission information was divided into three levels similar to Park and Jung's (2006) study:

> *Component* level: The knowledge needed for one operation step is related to the status and/or manipulation of an individual component. For example, mission information of the action "Open Valve 2" is one single operation related to only one function component in the spacecraft.
>
> *Subsystem* level: The knowledge needed for the required action is related to the statuses and/or manipulation of two or more components. For example, the action "Check if O_2 pressure $<$10 kPa AND the cabin pressure is between 80 kPa and 90 kPa" involves two components: oxygen and cabin pressure instruments.
>
> *System* level: The knowledge needed for the required action is related to a series of continuous state changes of two or more subsystems. For example, the action "Check the change of O_2 pressure AND the whole cabin pressure" involves system level information.

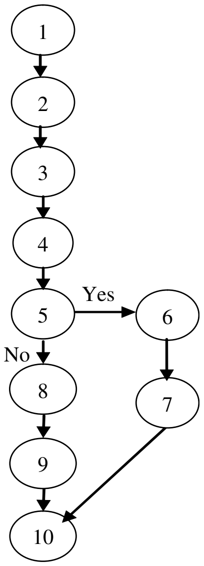


Fig. 4. Operation control graph of the example.

A mission information hierarchy graph was also developed as a data structure graph. Fig. 6 shows the knowledge levels of the seven steps. Here S is the root node. The nodes on the second level represent that there are seven functional steps in this operation unit (as shown in Fig. 3), and the downmost nodes represent that all the steps can be decomposed to the component level. The intermediate nodes denote the actions (information) of the corresponding steps.

The complexity of the operation step size is calculated from the second-order entropy of the operation control graph. For the above example shown in Fig. 4,

$$\mathrm{COSS} = H^2_{\mathrm{OCG}} = -\left(10 \times \frac{1}{10}\log_2\frac{1}{10}\right) = 3.322.$$

The complexity of the operation logic structure is calculated from the first-order entropy of the operation control graph. For the example shown in Fig. 4,

$$\mathrm{COLS} = H^1_{\mathrm{OCG}} = -\left(3 \times \frac{1}{10}\log_2\frac{1}{10} + \frac{7}{10}\log_2\frac{7}{10}\right) = 1.357.$$

The complexity of the operation instrument information is calculated from the second-order entropy of the operation instrument information graph. For the example shown in Fig. 5,

$$\mathrm{COII} = H^2_{\mathrm{OIIG}} = -\left(10 \times \frac{1}{14}\log_2\frac{1}{14} + 2 \times \frac{2}{14}\log_2\frac{2}{14}\right) = 3.236.$$

The complexity of the space mission information is calculated from second-order entropy of the mission information hierarchy graph. For the example shown in Fig. 6,

$$\mathrm{CSMI} = H^2_{\mathrm{MIHG}} = -\left(14 \times \frac{1}{18}\log_2\frac{1}{18} + \frac{4}{18}\log_2\frac{4}{18}\right) = 3.725.$$

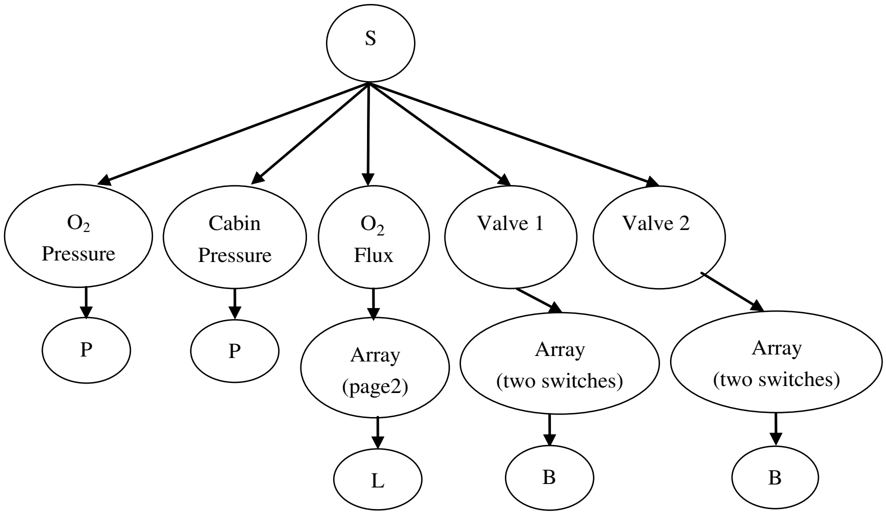


Fig. 5. Operation instrument information graph of the example.

Then the operation complexity of the example can be determined:

$$OC = \sqrt{(0.1725 \times COSS)^2 + (0.3821 \times COLS)^2 + (0.1965 \times COII)^2 + (0.2487 \times CSMI)^2} = 1.364.$$

2.2. Validation experiment

Since the complexity measure will be used for design evaluation and training optimization, it is reasonable to validate it by correlation analysis between complexity and performance data from training or specially designed experiments. In Park's study, the step complexity measure has been proven in evaluating the complexity of emergence operation procedure in nuclear plants (Park et al., 2001b, 2002). Because training data may contain the effects of various uncontrolled factors, in this study, the proposed complexity measure was validated by a one-factor experiment.

2.2.1. Subjects

It may be better to choose current astronauts as subjects. Unfortunately, the number and the free time of Chinese astronauts are very limited. It was impossible to have Chinese astronauts participate in this experiment. In the future, some new astronauts will be selected from among scientists and engineers. Thus 10 male engineers and researchers from China Astronaut Research and Training Center were recruited as subjects for this experiment. They participated in the experiment based on signed consent forms. Their ages ranged from 22 to 35 years old. They all had college education or above in science or engineering. These backgrounds are close to the standard for future astronaut selection.

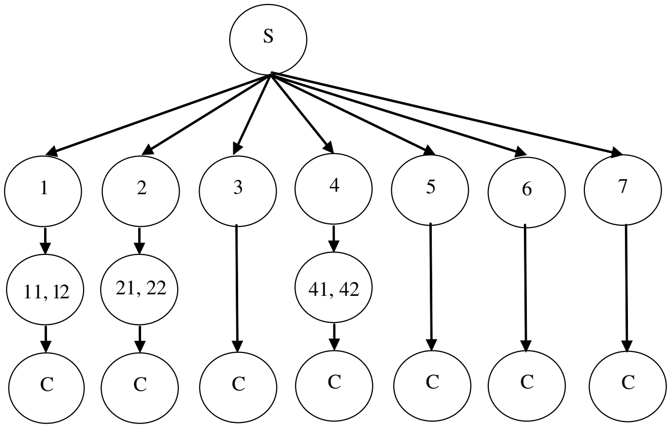


Fig. 6. Mission information hierarchy graph of the example.

2.2.2. Experiment design

The experiment design is shown in Fig. 7.

In this experiment, the independent variable was operation complexity. Seven operation units were chosen as the experiment tasks, and their operation complexity values ranged from 0.8 to 1.7.

Both objective and subjective indexes were used to evaluate the influence of operation complexity. The objective indexes included: (1) *Operation time*: the time in which the subject succeeded in finishing an operation unit; (2) *Error rate*: the total number of error trials divided by the total number of repeated trials. The subjective indexes included: (1) *Workload*: evaluated by the NASA Task Load Index (NASA-TLX) questionnaire; (2) *Subjective Complexity Rating (SCR)*: the perceived complexity rated by the subjects.

NASA-TLX is a multi-dimensional rating procedure based on six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration level (NASA Ames Research Center, 1987; Park and Cha, 1998; Noyes and Bruneau, 2007). Every dimension was divided into 10 scales. The workload value was obtained by adding up all the values of the six dimensions.

SCR is a single-item complexity rating, which was answered on a 1–10 scale (1 – extremely simple, 10 – extremely complex) during the last trial of the experiment by the subjects according to their own complexity perception. Such kind of single-item rating has been widely used to collect perceived complexity data from subjects because it is simple, straightforward, and proved to have good validity (e.g. Dong et al., 2007), although there are some other

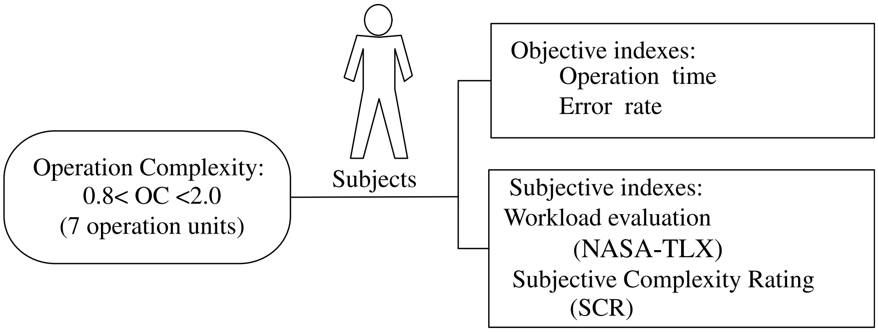


Fig. 7. Experiment design.

complicated methods such as the information complexity questionnaires for air traffic control displays (Xing, 2008).

As a subjective evaluation, the reliability and validity factors of SCR must be controlled (Murphy and Davidshofer, 2006). Then, an appropriate statistic might be chosen to summarize the degree of agreement between raters (Barrett, 2001). In this study, the description of the SCR rating was checked by several astronaut trainers. Before the formal experiment, the subjects were well trained and familiar with the seven operation units. Intraclass Correlation Coefficient (ICC, Shrout and Fleiss, 1979; McGraw and Wong, 1996; Barrett, 2001) was chosen to examine the reliability of the single-item rating. The ICC (A, 1) (McGraw and Wong, 1996) or ICC (2, 1) (Shrout and Fleiss, 1979) model was adopted in the examination.

2.2.3. Experiment tasks

The experiment tasks included seven spaceflight operation units with complexity values ranging from 0.8 to 1.7: monitoring operation, electrical power operation, thermal control operation, propulsion operation, guidance, navigation and control operation, environmental control operation, and overall checkout operation.

Each of the 10 subjects completed the seven operation units in random sequence. A subject finished a trial when he completed all the seven operation units. All subjects completed 18 trials except one subject, who completed only 17 trials. In all, the 10 subjects completed 179 trials.

2.2.4. Experiment facilities

The experiment platform was a full-scope spacecraft instrument system, which included the spacecraft instrument panel and the software modules (training support system). The instrument panel was composed of two monitors and eight types of switches, the same as those in a real spacecraft. Six types of switches were involved in this experiment. The training support system could act as the experiment support, which recorded important operation actions and the corresponding times. Another time-recording system was developed for this experiment to record the time required to perform every operation unit. A video-recording system was also installed in the experiment room.

2.2.5. Experiment procedure

The experiment was performed in the Instrument Panel Training Lab at the China Astronaut Research and Training Center. The experiment procedure included one preparation stage and two formal experiment stages.

The subjects were prepared according to the following steps: (1) read and listen to experiment introductions and sign a consent form signed; (2) fill in the required personal information; (3) listen to the experiment explanation and read detailed experiment instructions; (4) take a series of cognitive ability tests (not included in this paper); and (5) receive a theoretical training of the seven operation units with the operation handbook.

The first formal experiment stage included the following steps: (1.1) sign into the experiment; (1.2) receive theoretical training again, this time at the experiment platform; (1.3) single switch practice; (1.4) practice of the whole experiment process (the same as the formal experiment sequence) once; (1.5) perform a formal experiment for four trials; (1.6) take a break; (1.7) perform formal experiment for four more trials; and (1.8) fill in the questionnaire about the training methods and effects (not reported in this paper). This stage was mainly a learning stage.

The interval between the first and second formal experiment stages was three days. The second formal experiment stage included four steps: (2.1) perform the formal experiment for five trials; (2.2) take a break; (2.3) perform the formal experiment for five more trials; and (2.4) fill in the subjective questionnaire of NASA-TLX and complexity rating during the last trial. The first three trials were designed to allow the subjects to recover the skills acquired in the learning stage, and the performance during the 4th–9th trials at this stage will be analyzed for complexity measure validation.

By the two-stage formal experiment design, the learning process and operational skill recession can be observed (not reported in this paper). In addition, because each trial typically took 10–15 min, the fatigue effect could be avoided by the above procedure.

2.2.6. Data analysis

SPSS 13.0 for Windows™ was used to process the experiment data. Regression analysis was conducted to develop the quantitative relationship between the operation performance (operation time and error rate) and the operation complexity and that between the subjective evaluation/rating and the operation complexity. The fitness had little difference among different curve models. Thus linear models were chosen. The reliability analysis of SCR was examined with the Reliability Analysis function of SPSS, in which Intraclass Correlation Coefficient, Two-Way Random Model and Absolute Agreement were adopted.

3. Results

3.1. Complexity of selected operation units

The complexity values of the seven selected operation units described in Section 2.2.3 were calculated according to the method described in Section 2.1. The results are listed in Table 3.

3.2. Operation performance

Operation times and error rates of the 4th–9th trials at the second formal experiment stage were calculated and are listed in Table 4. The data were used for the validation of the proposed operation complexity measure.

3.3. Subjective evaluation

Results of subjective evaluation and rating for the seven selected operation units are listed in Table 5. The ICC value for SCR was 0.750, showing a good internal consistency among the subjects.

4. Discussion

4.1. Operation time

As shown in Fig. 8, the average operation time data were well proportional to the operation complexity values ($R = 0.876$, $p < 0.001$). The coefficients passed the *t*-test (both $p < 0.001$). Thus the operation complexity measure can predict the operation time quite well.

4.2. Error rate

Similarly, the error rates during the 4th–9th trials were analyzed, as shown in Fig. 9. Only the linear model was valid according to ANOVA test ($p = 0.004$). The coefficients passed the *t*-test ($p = 0.004$ and $p = 0.035$, respectively). Although the model is significant, the fitness was not satisfactory ($R = 0.343$). This implies that the error rate can only be partly predicted from operation complexity.

As described in the study of Morphew (2001), human errors result from various factors like poor cockpit design, stress-related factors, and training. Though the influence of control interfaces was considered in the operation complexity measure, there were other physiological and psychological factors that may have influenced the error rate, such as carelessness, fatigue, hastiness and many other factors addressed as performance shaping factors in the existing theory on human reliability or human error (Hollnagel, 1998; Sharit, 2006; Park, 1987). However, these factors are mostly not related to the interfaces and operation themselves and should not be included in the complexity measure. These factors are much more difficult to control during a validation experiment. Thus complexity measures often have limitations in the prediction of error rate. This has been found in other studies (e.g. Xu et al., 2009). Therefore, error rate is not considered (Park et al., 2002) or considered as a secondary criterion for similar validation.

Table 3
Operation complexity values of the selected operation units.

Operation unit	Operation Complexity				
	COSS	COLS	COII	CSMI	Weighted Sum
1. Monitoring operation	1.585	1.585	2.000	2.128	0.936
2. Electrical power operation	2.000	2.000	2.585	2.281	1.133
3. Thermal control operation	3.000	2.156	2.750	3.107	1.355
4. Propulsion operation	3.322	1.722	3.624	2.810	1.325
5. Guidance, navigation, and control operation	3.459	2.005	3.975	2.869	1.436
6. Environmental control operation	3.807	1.921	4.004	3.866	1.585
7. Overall checkout operation	4.248	1.890	4.806	3.361	1.627

Note: COSS – complexity of operation step size; COLS – complexity of operation logic structure; COII – complexity of operation instrument information; and CSMI – complexity of space mission information.

Table 4
Operation performance of the selected operation units.

Operation unit	Second formal experiment stage (4th–9th trials)			
	Operation time (s)		Error rate	
	M	SD	M	SD
1. Monitoring operation	5.80	1.27	0.0167	0.0527
2. Electrical power operation	10.9	2.62	0.0167	0.0527
3. Thermal control operation	22.5	5.36	0.0333	0.0703
4. Propulsion operation	47.8	7.40	0.0333	0.0703
5. Guidance, navigation, and control operation	37.5	6.78	0.0167	0.0527
6. Environmental control operation	81.3	12.74	0.17	0.112
7. Overall checkout operation	83.5	11.57	0.05	0.0805

Moreover, in the experiment, error rates were calculated from a limited number of trials. The calculation of the error rate then had a lower accuracy and may also have contributed to the bad correlation between operation complexity and error rate.

4.3. Subjective evaluation

The curve estimation of NASA-TLX scores and operation complexity values is shown in Fig. 10. The regression result indicates that the NASA-TLX scores increased with operation complexity values statistically ($R = 0.698$). The ANOVA results show that the model was statistically significant ($p < 0.001$). The coefficients passed the *t*-test (both $p < 0.001$).

The SCR and operation complexity scores were also modeled. Fig. 11 shows the regression analysis and ANOVA results. It can be seen that the model was statistically significant ($p < 0.001$). The coefficients passed the *t*-test (both $p < 0.001$). The fitness was higher ($R = 0.802$) than that of the NASA-TLX ($R = 0.698$).

From the above two models, it can be concluded that the operation complexity measure was reasonably consistent with subjective evaluation and rating.

The relation of the operation complexity and operation performance/workload has been studied in other areas. It was proven in the nuclear power plant operation that high-complexity tasks and lack of training may lead to a longer operation time, higher error rate and subjective workload (Xu et al., 2008). Though the mental workload is a multi-dimensional and meditational construct of many factors in aviation (Gopher and Donchin, 1986; Hart and Staveland, 1988), Hancock's study indicated that the subjective workload scores are sensitive to task difficulty (Hancock et al., 1995). Svensson and Angelborg-Thanderz (1997) also found that pilot mental workload was affected by task difficulty. Particularly, the statistical analysis of the data collected in long-term spaceflights has revealed the relationship between astronaut errors frequency and task complexity (Nechaev et al., 1998). Of course, the above studies do not mean that human error should be blamed on

Table 5
Subjective evaluation and rating results.

Operation unit	NASA-TLX		SCR	
	M	SD	M	SD
1. Monitoring operation	8.80	2.74	1.10	0.32
2. Electrical power operation	10.5	3.66	1.60	0.52
3. Thermal control operation	13.6	4.67	2.80	1.14
4. Propulsion operation	21.5	8.10	4.60	1.96
5. Guidance, navigation and control operation	20.2	7.27	3.90	0.57
6. Environmental control operation	27.6	8.15	6.60	1.78
7. Overall checkout operation	28.2	8.98	7.00	1.77

Note: SCR – Subjective Complexity Rating.

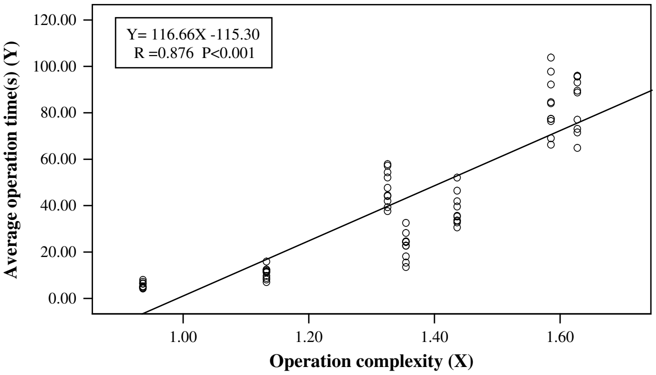


Fig. 8. Linear regression of average operation time.

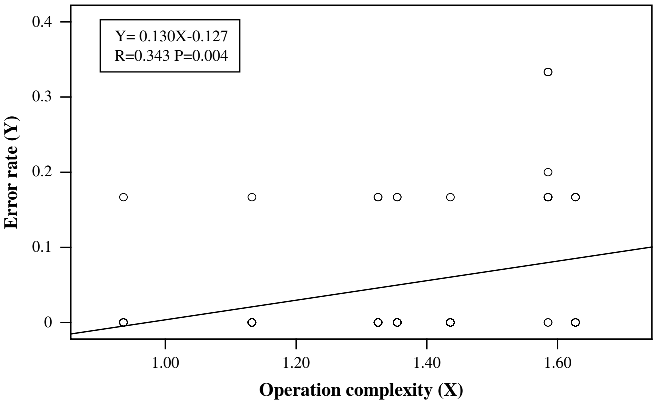


Fig. 9. Linear regression curve of error rate.

task complexity solely; rather, that human error in aviation is a result of factors like poor cockpit design, stress-related factors and training (Morphew, 2001). The results of this study seem to be consistent with the other studies, and by reflecting such relationships, the proposed complexity measure can be used for performance prediction, design evaluation, and training arrangement reference.

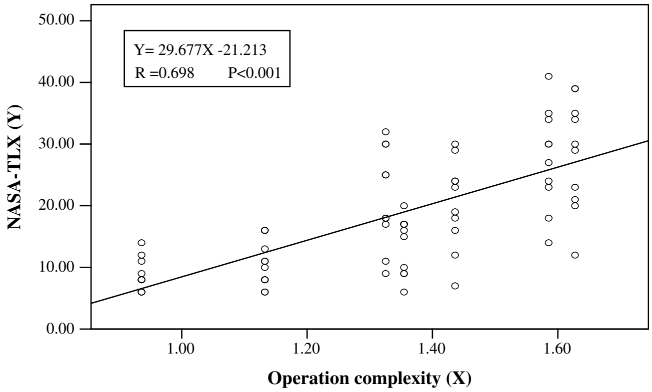


Fig. 10. Linear regression curve of NASA-TLX.

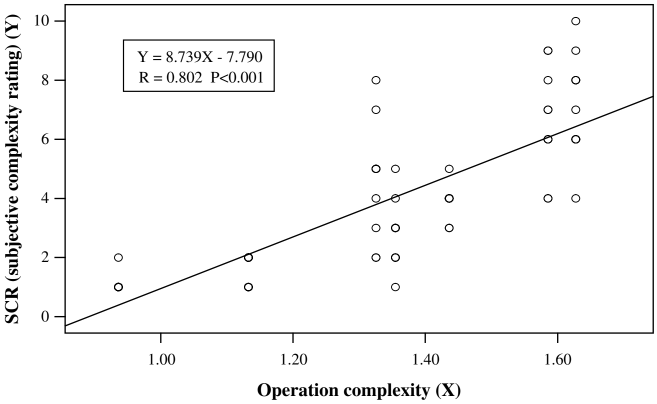


Fig. 11. The linear regression curve for subjective complexity rating.

5. Conclusions

A one-factor experiment was designed in this study to verify the suitability and validity of a proposed operation complexity measure for spaceflight. Ten subjects participated in this experiment, which included seven operation units. The operation performance and subjective evaluation were compared with the operation complexity values. The regression analysis results show that the average operation time and subjective complexity rating can be predicted well from the operation complexity, and the workload can also be estimated from the operation complexity quite well; however, the error rate can only be partly explained by the operation complexity. This might be because the proposed operation complexity measure does not reflect other factors influencing human error in spaceflight operations.

The operation complexity measure has been applied to astronaut training in recent missions in China and it has been demonstrated as being suitable in evaluating a design for training and test of spaceflight operations.

There are other issues to be considered in the operation complexity measure in the future. The weights of the four factors need to be further verified by more spaceflight experts and through experiment data analysis. Moreover, astronaut training data will be acquired and analyzed to help in the determination of the weights. The classification of the instrument types and knowledge levels may need to be further discussed by experts.

There were some limitations in the experiment. The experiment system was one full-scope control panel system in the astronaut training center without space environment effects, such as microgravity, a small cabin or flight suits. The subjects participating in the experiment were researchers and engineers rather than astronauts.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (70771056), the Manned Space Project of China, and the Program for New Century Excellent Talents in University.

References

Barrett, P., 2001. Assessing the reliability of rating data. http://www.pbarrett.net/techpapers/rater.pdf (accessed 02.02.09).

Davis, J.S., LeBlanc, R.J., 1988. A study of applicability of complexity measures. IEEE Transactions on Software Engineering 14 (9), 1366–1372.

Dong, Y.F., Ling, C., Hua, L.S., 2007. Effect of glance duration on perceived complexity and segmentation of user interfaces. In: Jacko, J.A. (Ed.), 12th International

Conference on Human–Computer Interaction (HCII2007), Beijing, China. HCI Intelligent Multimodal Interaction Environments, pp. 605–614.

Edmonds, B., 1999. What is complexity? – the philosophy of complexity *per se* with application to some examples in evolution. In: Heylighen, F., Aerts, D. (Eds.), The Evolution of Complexity, pp. 1–18.

Gopher, D., Donchin, E., 1986. Workload: an examination of the concept. In: Boff, K.R., Kaufman, L., Thomas, J.P. (Eds.), Handbook of Perception and Performance. Cognitive Processes and Performance, vol. 2. Wiley, New York, NY.

Gonzalez, R.R., 1995. A unified metric of software complexity: measuring productivity, quality, and value. Journal of Systems and Software 29, 17–37.

Hancock, P.A., Williams, G., Manning, C.M., Miyake, S., 1995. Influence of task demand characteristics on workload and performance. International Journal of Aviation Psychology 5, 63–86.

Hart, S.G., Staveland, L.E., 1988. Development of a multi-dimensional workload rating scale: results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. Elsevier, Amsterdam, The Netherlands.

Hollnagel, E., 1998. Cognitive Reliability and Error Analysis Method (CREAM). Elsevier Science Ltd, Amsterdam.

Ji, Y.M., Zhang, J.L., 2005. Determining the weights of evaluation indexes of supply chain performance with FAHP. Journal of Central China Normal University (Natural Sciences) 39 (2), 190–194.

Lew, K.S., Dillon, T.S., FoMrard, K.E., 1988. Software complexity and its impact on software reliability. IEEE Transactions on Software Engineering 14 (1), 1645–1655.

McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1 (1), 30–46.

Mikhailov, L., Tsvetinov, P., 2004. Evaluation of services using a fuzzy analytic hierarchy process. Applied Soft Computing 5 (1), 23–33.

Morphew, M.E., 2001. Psychological and human factors in long duration spaceflight. McGill Journal of Medicine 6 (1), 74–80.

Mowshowitz, A., 1967. Entropy and the Complexity of Graphs. Ph.D. Dissertation. University of Michigan, pp. 101–113.

Murphy, K.R., Davidshofer, C.O., 2006. Psychological Testing: Principles and Applications (translated by Zhang, N., Yang, Y.S., Xu, A.H.), sixth ed. Shanghai Academy of Social Science Press, pp. 102–177 (in Chinese).

NASA Ames Research Center, 1987. Manual of NASA Task Load Index (TLX). v. 1.0 (Computerized Version). NASA Ames Research Center.

Nechaev, A.P., Myasnikov, V.I., Stepanova, S.I., Kozerenko, O.P., Isaev, G.F., Bronnikov, S.V., 1998. Methodological approach to study of cosmonauts errors and its instrumental support. Acta Astronautica 42 (1–8), 331–338.

Nelson, W.R., Haney, L.N., Ostrom, L.T., Richards, R.E., 1998. Structured methods for identifying and correcting potential human errors in space operations. Acta Astronautica 43, 211–222.

Nelson, W.R., 1999. Applying lessons learned to enhance human performance and reduce human error for ISS operations. AIP Conference Proceedings 458, 420–425.

Noyes, J.M., Bruneau, D.P.J., 2007. A self-analysis of the NASA-TLX workload measure. Ergonomics 50 (4), 514–519.

Ostrom, L.T., Gertman, D.I., Blackman, H.S., 1992. Assessing human reliability in space – what is known, what still is needed. In: Proceedings of the Space Programs and Technologies Conference, Huntsville.

Park, J., Jung, W., Ha, J., 2001a. Development of the step complexity measure for emergency operating procedures using entropy concepts. Reliability Engineering and System Safety 71 (2), 115–130.

Park, J., Jung, W., Kim, J., Ha, J., Shin, Y., 2001b. The step complexity measure for emergency operating procedures – comparing with simulation data. Reliability Engineering and System Safety 74, 63–74.

Park, J., Jung, W., Ha, J., Park, C., 2002. The step complexity measure for emergency operating procedures: measure verification. Reliability Engineering and System Safety 76, 45–59.

Park, J., Jung, W., 2006. A study on the development of a task complexity measure for emergency operating procedures of nuclear power plants. Reliability Engineering and System Safety 92 (8), 1102–1116.

Park, K.S., 1987. Human Reliability: Analysis, Prediction, and Prevention of Human Errors. Elsevier Science Publishing, Amsterdam.

Park, P., Cha, D., 1998. Comparison of subjective mental workload assessment techniques for the evaluation of in-vehicle navigation system usability. In: Proceedings of Session T56 of the 5th World Congress of ITS 98.

Pierre, J.T., Gregory, J.H., 1995. Extravehicular activity training and hardware design consideration. Acta Astronautica 36 (1), 13–26.

Rauterberg, M., 1992. A method of a quantitative measurement of cognitive complexity. In: Van der Veer, G.C., Tauber, M., Bagnara, S., Antalovits, M. (Eds.), Human–Computer Interaction: Tasks and Organisation. CUD Publishers, Roma, pp. 295–307.

Saaty, T.L., 1980. The Analytic Hierarchy Process. McGraw-Hill, New York.

Saaty, T.L., 1988. Multicriteria Decision Making: The Analytic Hierarchy Process. RWS Publications, Pittsburgh, PA.

Seastrom, J.W., Peercy, R.L., Johnson, G.W., Sotnikov, B.J., Brukhanov, N., 2004. Risk management in international manned space program operations. Acta Astronautica 54 (4), 273–279.

Sharit, J., 2006. Human error. In: Salvendy, G. (Ed.), Handbook of Human Factors and Ergonomics, third ed. John Wiley & Sons, Hoboken, pp. 708–712.

Shayler, D.J., 2005. Disasters and Accidents in Manned Spaceflight (translated by Yuan, J., & Zhen, M.). China Astronautics Press, Beijing, p. 265–282 (in Chinese).

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 86 (2), 420–428.

Svensson, E., Angelborg-Thanderz, M., 1997. Information complexity – mental workload and performance in combat aircraft. Ergonomics 40 (3), 362–380.

Tang, Y., Beynon, M.J., 2005. Application and development of a fuzzy analytic hierarchy process within a capital investment study. Journal of Economics and Management 1 (2), 207–230.

Wang, D.H., Huang, W.F., 1996. Human reliability in manned spaceflight. Space Medicine & Medical Engineering 9 (4), 295–301 (in Chinese).

Wang, Z., Wang, L., Xu, R.T., 2006. Study on operational reliability of astronaut in spacecraft cabin. Space Medicine & Medical Engineering 19 (5), 358–362 (in Chinese).

Wei, S.X., Zhou, X.Z., 1998. Multi-Attribute Decision-Making Theory and the Application in the C3I System. Defense Industry Press, p. 173–176 (in Chinese).

Xing, J., 2004. Measures of Information Complexity and the Implications for Automation Design. Federal Aviation Administration, Washington DC. Report No: DOT/FAA/AM-04/17.

Xing, J., Manning, C., 2005. Complexity and Automation Displays of Air Traffic Control: Literature Review and Analysis. Federal Aviation Administration, Washington DC. Report No: DOT/FAA/AM05-4.

Xing, J., 2008. Designing Questionnaires for Controlling and Managing Information Complexity in Visual Displays. Federal Aviation Administration, Washington DC. Report No: DOT/FAA/AM-08/18.

Xu, S., Song, F., Li, Z.Z., Zhao, Q.Y., Luo, W., He, X.H., Salvendy, G., 2008. An ergonomics study of computerized emergency operating procedures: presentation style, task complexity, and training level. Reliability Engineering and System Safety 93 (10), 1500–1511.

Xu, S., Li, Z.Z., Song, F., Luo, W., Zhao, Q.Y., Salvendy, G., 2009. Influence of step complexity and presentation style on step performance of computerized emergency operating procedures. Reliability Engineering and System Safety 94 (2), 670–674.

Zhang, Y.J., Wu, B., Li, Z.Z., Cao, X.S., Xu, S., 2007. Operation complexity measure of emergency failure operation procedure in spaceflight. In: Proceedings of the 16th IAA Humans in Space Symposium, Beijing.

Zhang, Y.J., Wu, B., Wang, Y., Shi, L.J., Li, Z.Z., Wu, S., 2008. Determination of index weights in the operation complexity model of spaceflight. Space Medicine & Medical Engineering 21 (3), 252–256 (in Chinese).