Dynamic modification of the G Rasch’s logistic model for the tasks of examination the engineers’ knowledge

A Kirillin
Technology of testing and operation Department, Moscow Aviation Institute (National Research University), 4 Volokolamskoe shosse, Moscow 125993, Russian Federation
E-mail: kirillinav@mati.ru

Abstract. Problem of efficient and safe operation of complex technical systems is always one of the most important in its life cycle. Analysis of the failures and emergencies statistics shows a trend in the growth of the human factor and a decrease of failures for technical reasons. Unfortunately, it is not possible to completely exclude human from the control system due to adherence to safety standards and industry requirements. However, issues of evaluating the professional suitability of personnel operating complex equipment, as well as issues of organizing and planning the training process, are becoming relevant. Based on a comparison of the statistical model of Georg Rasch, used to assess the preparedness of students and the degree of complexity of control tasks, with the dynamic universal model of reliability growth in the development process, used in various practical applications, a methodology for the analysis and planning of training effectiveness is developed. The algorithm of operation and application of the described methodology is illustrated by a number of examples. The dynamic modification of the presented Rasch’s statistical model allows predicting the degree of preparedness of homogeneous groups of students, thereby making it possible to plan the number of personnel training cycles for admission to work with complex and potentially dangerous technical systems.

1. Introduction
The reliability and safe operation of complex technical systems cannot be evaluated from the point of only the technical part. Since in this case it is assumed that the performance characteristics of the operating engineering personnel are accepted as optimal and there is no risk of error. Obviously, such a situation cannot arise in actual operation. For complex technical systems, in which human is an integral part of the control loop, there are two options for responding to a possible violation of the operating mode: the first one is automatic, when signs of violation appear; the second – duplicate manual control carried out by an engineer in case of violation of the operating mode according to light, sound or other indications. At the same time, a decrease in the inertia of the engineer-operator reaction, as well as the accuracy of his actions, significantly affects the negative consequences of the possible failures. Therefore, there is a need to develop various methodological methods for evaluation and ensuring the reliability of the ergonomic component of the system which is an engineer-operator.

Currently there are many approaches to evaluate reliability of human, as part of a technical system. Moreover most of them are the basis of the HRA methods (Human Reliability Analysis) and are well described by E Hollnagel [1] and H Liao [2-4]. In engineering practice the applicability of various HRA methods is considered. For example, Zhang Li [5] considers approaches to evaluate the human cognitive reliability (HCR) as an addition to the standard THERP procedure (The Technique for Human Error-Rate Prediction) for Qinshan NPP staff. Engineers of University in Brno (Czech Republic) applied the
Monte Carlo method to calculate the most important indicator of the human reliability working in the control loop of a complex technical system – the Mean Time Between Failures (MTBF) [6].

Analysis of the publications showed that HRA methods and models evaluate reliability either from the position of a probabilistic analysis of errors or from the position of cognitive reliability. At the same time, an engineer is considered as a specialist having 100% knowledge about the facility being operated, the rules and methods of working with it. In practice this situation is far from reality. Therefore, the author sets the task of developing a methodology for evaluation of the level of competence and knowledge of engineer-operator, which in combination with the HRA methods application would provide more complete information about the reliability of human as an element of a complex technical system.

2. Methodology and research methods

2.1. G Rasch’s logistic model

The simplest and most common method of assessing the success is test control estimating the answers to control questions in a nominal scale (right - wrong) and an integral assessment of a group of students – an evaluation of the probability of success by frequency $R = \frac{M}{N}$, where $M$ – the number of right answers, $N = k \cdot n$ – selection size, $n$ – the number of students in the group, $k$ – the number of control questions [7].

This probability, which characterizes the assimilation degree of the taught material, depends both on the initial preparedness of the students and on the content of the taught discipline and the professional skill of the teacher. The level of complexity of the material should be correlated with the degree of preparedness of students and the perception of this material [8-14]. A statistical model that allows to evaluate the effectiveness of training, taking into account both the degree of preparedness of students and the degree of complexity of the taught material, is the logistic model of Georg Rasch [15]:

$$P(\theta - \beta) = \frac{\exp((\theta - \beta))}{1 + \exp((\theta - \beta))}$$ (1)

where $\beta$ – level of difficulty of control task, $\theta$ – level of preparedness of students.

Logistic probability distribution $P(\theta - \beta)$ is symmetrical with zero expected value (at $\theta = \beta$, $P(\theta - \beta) = 0.5$ ), dispersion $\pi^2/3$ and $\beta^2 = 4.2$.

The logistic distribution in the middle part is close to normal. However, it has more “heavy tails”, i.e. the probabilities of both small and large values of a random variable $(\theta - \beta)$ are greater than that of the normal law of probability distribution ($\beta^2 > 3$ for a normal distribution and for a logistic distribution it is more peaked).

The logistic distribution is simpler than normal because the calculation of the function values of this distribution does not require the use of special tables or programs.

To carry out the calculations using the logistic distribution of Rasch, the transformation

$$\ln \frac{p}{q} = \theta - \beta$$

where $q = 1 - p$.

When organizing an active experiment it is possible to evaluate the preparedness degree of a group of students with a fixed complexity degree as well as the complexity degree with a fixed group composition [16-18].

The Rasch’s model is statistical, i.e. it allows a one-time evaluation of the training effectiveness.

2.2. Analysis of G Rasch’s model from the perspective of the general reliability theory

When preparing operators for working with complex equipment, the reliability and operational safety of the equipment depend on the human factor reliability. Therefore, the requirements for the
preparedness of the human operator are extremely high. In this case one-time training may not be enough. The question of planning the entire training process arises, i.e. firstly the question of the number of training cycles.

This problem can be solved using well-known models of reliability growth. These models are widely used in variable areas as development models, for example, in substantiating the number of test cycles during experimental testing of aircraft [19]. The task of evaluating the dynamics of training effectiveness close to the task of the dynamics reliability assessment of the product, depending on the number of test cycles. After each cycle the product is refined in order to eliminate defects identified during testing, increasing its reliability.

Indeed, the control of students' knowledge can be considered as their test, and the identified undeveloped fragments, as learning defects that require more detailed study.

3. Results and discussion

3.1. Dynamic modification of the G Rasch’s logistic model

One of the models of reliability growth is the logistic model [20]

$$ R(n) = \frac{R_0}{R_0 + (1 - R_0) \exp(-\Omega n)} $$

(2)

At the number of cycles \( n = 0 \), \( R = R_0 \); at \( n \to \infty \), \( R \to 1 \).

It is easy to see that model (2) is a special case of model (1), when \( R_0 = 0.5 \), \( n = 1 \), \( \Omega = \theta - \beta \) and only positive values \( \Omega \) are considered leading to increased reliability \( R \). Also in the model (2), the parameter \( \Omega = \text{const} \), but in the model (1) \( \Omega = \theta - \beta \) – random variable, the indirect measurement of which is based on an estimate of the probability \( P \) by the frequency.

Model (2) can be written in the equivalent discrete recurrent form

$$ R_i = \frac{R_{i-1} \exp(\Omega)}{R_{i-1} \exp(\Omega) + (1 - R_{i-1})} $$

whence after simple transformations the following is obtained

$$ \ln \frac{R_i}{q_i} = \ln \frac{R_{i-1}}{q_{i-1}} + \Omega, $$

and finally passing to the designation of model (1), the prediction equation for 1 step is obtained:

$$ \ln \frac{p_i}{q_i} = \ln \frac{p_{i-1}}{q_{i-1}} + (\theta_{i-1} - \beta_{i-1}) $$

(3)

where \( (\theta_{i-1} - \beta_{i-1}) \) – the estimates obtained by processing information on previous cycles up to \( (i - 1) \).

It should be noted that there is the simplest linear model of the dependence of the growth rate \( \Omega_i \) on the number of cycles is used:

$$ \Omega_i = (\theta - \beta) \eta_i. $$

It is rational to use the arithmetic mean estimate as estimate of a random variable \( (\theta_{i-1} - \beta_{i-1}) \)

$$ \Omega_{i-1} = \frac{1}{n-1} \sum_{j=1}^{n-1} \ln \frac{p_j}{q_j}. $$

Equation (3) can be expanded to predict more than one step forward, for example:

$$ \ln \frac{p_{i-1}}{q_{i-1}} = \ln \frac{p_{i-1}}{q_{i-1}} + (\theta_{i-1} - \beta_{i-1}) \cdot 2 $$

eq 3
3.2. Examples of using dynamic modification of the G Rasch’s model

Example 1. The set value of $P_{set} = 0.99$, i.e. $\ln \frac{0.99}{0.01} = 4.595$. Then the value $\ln \frac{p_1}{q_1}$, obtained after the first familiarization control of knowledge is:

$$\ln \frac{p_1}{q_1} = \ln \frac{0.6}{0.4} = 0.405 = \theta_i - \beta_i,$$

as in (3) $\ln \frac{p_2}{q_2} = \ln \frac{0.5}{0.5} = 0$.

Then, after the second training cycle the following values are expected to be reached:

$$\ln \frac{p_2}{q_2} = 0.405 \cdot 2, \quad \ln \frac{p_2}{q_2} = 2.25,$$

whence $P_2 = 0.692$.

With such reliability growth rate, for reaching the value $P_{set} = 0.99$ it will be required $4.595 - 0.405 = 4.05 \cdot n$ training cycles, whence $n \approx 10$.

Example 2. There is the preparedness of the group of students is higher, $p_i = 0.8$. Then $\ln \frac{0.8}{0.2} = 1.386$,

and for reaching the value $P_{set} = 0.99$ it will be already required $n \approx 3$ training cycles.

Example 3. The set value $P_{set}$ is very high, and equal to 0.999999, i.e. $\ln \frac{0.999999}{0.000001} = 13.8$. At the level of preparedness $p_i = 0.8$, in this case $n = \frac{13.8 - 1.386}{1.386} \approx 9$ training cycles will be required.

In the practical application of the proposed approach there is the question of the decisions reliability is raised. The probability in the model (1) is a monotonically increasing function of the value $(\theta - \beta)$. Thus, the lower confidence limit of the evaluation of this probability by frequency corresponds to the lower confidence limit of an indirect measurement $(\theta - \beta)$. The result is a simple practical technique for taking errors into account when replacing the true probability value $P$ with its estimate $\hat{P} = \frac{M}{N}$, shown below with a specific example.

Example 4. Assume that as a result of familiarization test of 10 students on 10 control questions ($N = n \cdot k = 10 \cdot 10 = 100$), there were received 80 correct answers, i.e. $\hat{P} = \frac{80}{100} = 0.8$.

Using the normal approximation of the binomial distribution, the lower 90% confidence limit is obtained at $\gamma = 0.9$:

$$P_L = \hat{P} - u_{0.9} \sqrt{\frac{\hat{P}(1 - \hat{P})}{n}} = \frac{0.8 \cdot 0.2}{100} \approx 0.75,$$

where $u_{0.9}$ – quantile of standard normal distribution.

Further calculation is carried out analogously to example 3, i.e. $\ln \frac{0.75}{0.25} = 1.099$, and the upper number of cycles is $n_H = \frac{13.8 - 1.099}{1.099} \approx 11$.

Herewith, the predicted value after the second training cycle will be $\ln \frac{p_2}{q_2} = 1.099 \cdot 2$, whence $P_2 = 0.9$.

Suppose further, that during testing on the second cycle, 90 correct answers were obtained, i.e. $\hat{P} = \frac{90}{100} = 0.9$. 

4
The lower confidence limit at $\gamma = 0.9$ will be:

$$P_L = 0.9 - 1.28 \sqrt{\frac{0.9 - 0.1}{100}} \approx 0.862\ ,$$

whence $\theta_2 - \beta = \ln \frac{0.862}{0.138} - \ln \frac{0.75}{0.25} = 1.832 - 1.099 = 0.733\ .$

The average growth rate will be $\frac{1.099 - 0.733}{2} = 0.916$ and the predicted number of cycles for reaching $P_L = 0.999999$ will be $n_H = \frac{13.8 - 1.832}{0.916} \approx 13$, i.e. estimation errors $\hat{P}$ already significantly affect the number of training cycles.

To reduce the required number of training cycles it is necessary to increase its effectiveness.

**Example 5.** Suppose that under the remaining conditions of Example 4, the number of correct answers after the second training cycle increased to 95. Then $\hat{P} = 0.95$, the lower confidence limit will be:

$$P_L = 0.95 - 1.28 \sqrt{\frac{0.95 - 0.05}{100}} \approx 0.922\ ,$$

whence $\theta_2 - \beta = \ln \frac{0.922}{0.078} - \ln \frac{0.75}{0.25} = 2.47 - 1.099 = 1.371\ .$

The average growth rate will be $\frac{1.099 - 1.371}{2} = 1.235\ , \text{ and } n_H = \frac{13.8 - 2.47}{1.235} \approx 9$ is required, i.e. an increase in the training effectiveness by 5% leads to a decrease in the training volume by more than 30%.

### 3.3. Discussion

The dynamic modification of the G. Rasch's logistic model can also be used in problems of comparing different groups of operators from the standpoint of preparedness. In contrast to traditional models of 2-way-ANOVA, which is the main mathematical model for solving such problems, the Rasch's model allows not only to assess the influence of complexity factors of control tasks and the preparedness of engineers, but also to calculate the probability of successful training in both discrete and continuous scales.

The dynamic modification of Rasch's model also allows forecasting and planning the number of training cycles, which is not possible in case using ANOVA models.

In contrast to the models of ANOVA, the presented mathematical models, in the opinion of the author of the report, look much simpler.

### 4. Conclusion

However, the above methods are very sensitive to the composition of students groups, as described in detail in [18], therefore the author does not recommend their use for dynamically recruited groups, the composition of which is constantly changing. Results of the analysis of these groups have very uncertain character. At the same time, the proposed technique showed itself very well in planning the training cycles for a group during a fixed period of time, for example, a semester. It can be recommended for use as an alternative to G Ebbinghaus's approach for planning the training cycles to ensure the validity of the results. Thus, the developed technique allows to plan and manage a multi-stage learning process, which ensures its effectiveness.

### References

[1] Hollnagel E 1996 Reliability Analysis and Operator Modelling. *Reliability Engineering and System Safety* 52 327 doi: 10.1016/0951-8320(95)00143-3
[2] Liao H, Forester J, Dang V N, Bye A, Chang Y H, Lois E 2019 Assessment of HRA method predictions against operating crew performance: Part I: Study background, design and methodology. Reliab. Eng. and System Safety 191 **106509** doi: 10.1016/j.ress.2019.106509

[3] Liao H, Forester J, Dang V N, Bye A, Chang Y H, Lois E 2019 Assessment of HRA method predictions against operating crew performance: Part II: Overall simulator data, HRA method predictions and intra-method comparison. Reliability Engineering and System Safety 191 **106510** doi: 10.1016/j.ress.2019.106510

[4] Liao H, Forester J, Dang V N, Bye A, Chang Y H, Lois E 2019 Assessment of HRA method predictions against operating crew performance: Part III: Conclusions and achievements. Reliability Engineering and System Safety 191 **106511** doi: 10.1016/j.ress.2019.106511

[5] Zang L, He X, Dai L, Huang X 2007 The Simulator experimental study on the Operator Reliability of Qinhan nuclear power plant. Reliability Engineering and System Safety 92 252 doi: 10.1016/j.ress.2005.12.005

[6] Jirgl M, Havlikova M, Bradac Z, Stastny L 2016 Monte Carlo Reliability Analysis of System with Human Operator. IFAC-PaperOnLine 49(25) 272 doi: 10.1016/j.ifacol.2016.12.046

[7] Avanesov V S 1980 The problem of psychological tests. Soviet Education 22(6) 6

[8] Donihue M, Mattos J, Theoharides C, Tiffany C 2006 Meeting the Standards: An Analysis of Educational Assessment Test Scores in Maine. Goldfarb Center Working Paper 2006-0003

[9] Haertel E 2018 Tests, Test Scores, and Constructs. Educational Psychologist 53 14 doi: 10.1080/00461520.2018.1476868

[10] Thissen D, Wainer H 2002 Test Scoring. Journal of Educational Measurement 39(3) 265 doi:10.1111/j.1745-3984.2002.tb01178.x

[11] Engelhardt L, Goldhammer F 2019 Validating Test Score Interpretations Using Time Information. Frontiers in Psychology 10 1131 doi:10.3389/fpsyg.2019.01131

[12] Kleinberg J, Raghu M 2015 Team Performance with Test Scores. Proceedings of the Sixteenth ACM Conference on Economics and Computation - EC’15 (Portland, OR, USA) doi:10.1145/2764468.2764496

[13] Rasch G E 1993 Probabilistic Models for Some Intelligence and Attainment Tests (Chicago, IL, USA)

[14] Peterson J K 2016 Logistics Models. Cognitive Science and Technology 257 doi:10.1007/978-981-287-874-8_13

[15] Hand D J, van der Linden W J, Hambleton R K 1998 Handbook of Modern Item Response Theory. Biometrics 54(4) 1680 doi:10.2307/2533698

[16] Aleksandrovskaya L N, Kirillin A V, Shumskaya L P 2014 Mathematical models in assessing the effectiveness of continuing education. Quality. Innovations. Education 8(111) 30 [in Russian]

[17] Aleksandrovskaya L N, Kirillin A V, Iosifov P A, Mitrofanova I P 2015 A model for evaluating the effectiveness of retraining of specialists in multilevel studying. Quality. Innovations. Education 1(116) 3 [in Russian]

[18] Aleksandrovskaya L N, Kirillin A V, Borisova E V 2018 Statistical analysis of homogeneity of students’ preparedness. Quality. Innovations. Education 6(157) 5 [in Russian]

[19] Aleksandrovskaya L N, Kruglov V I, Kuznetcov A G, Kuznetcov V A, Kutin A A, Sholom A M 2003 Theoretical basis of the test and experimental development of complex technical systems (Moscow: Logos)

[20] Tsoularis A, Wallace J 2002 Analysis of logistic growth models. Mathematical Biosciences 179(1) 21 doi:10.1016/s0025-5564(02)00096-2