Sensors & Transducers, Vol. 237, Issue 9-10, September-October 2019, pp. 165-170

Sensors & Transducers
Published by IFSA Publishing, S. L., 2019
http://www.sensorsportal.com

Evaluation Results of Testing of the Measuring Instruments Software

[1, *] O. VELYCHKO, [2] O. HRABOVSKYI and [2] T. GORDIYENKO
[1] State Enterprise 'All-Ukrainian State Scientific and Production Centre for Standardization, Metrology, Certification and Protection of Consumer', (SE 'Ukrmetrteststandard'), 4 Metrologichna str., Kyiv, 03143, Ukraine
[2] Odesa State Academy of Technical Regulation and Quality, 15 Kovalska str., Odesa, 65020, Ukraine
[1] Tel.: + 38 050 464 00 24, + 38 097 177 57 43, fax: + 38 044 526 03 35
[*] E-mail: velychko@hotmail.com

Received: 30 August 2019 /Accepted: 27 September 2019 /Published: 30 November 2019

Abstract: The developed special methodology of evaluation results of the testing of measuring instruments software (MI SW) with taking into account all essential requirements of OIML document and WELMEC guidelines was presented. This methodology contains the methodology for testing of MI SW on the following generalized indicators. The developed methodology for the testing of some varieties of MI SW with built-in computer was used. The obtained results showed the possibility of applying the Rasch model for the analysis of the scale for evaluating of results of the testing MI SW.

Keywords: Software, Measuring instruments, Testing, Evaluation, Rasch model.

1. Introduction

Specialized software (SW) for measuring instruments (MIs) is subject to testing for conformity assessment of MI. The testing of MI SW is devoted to the documents of the International Organization of Legal Metrology (OIML) [1] and the guidelines of the European Organization for Cooperation in Legal Metrology (WELMEC) [2-3].

Since 2005 in Ukraine, national metrology institutes have been carry out of work on the testing of MI SW [4-5]. In 2013, the national standard DSTU 7363 [6] has been developed, which sets the general technical requirements for MI SW. This standard only partly meets the requirements of international and regional documents and guidelines. On the basis of DSTU 7363 developed special methodical recommendations for the testing of SW MI.

In [7], a comparative analysis of the general requirements in the documents and guidelines of the international and regional organizations of legal metrology OIML and WELMEC concerning the testing of MI SW was conducted. In this work a universal algorithm for testing of MI SW is proposed.

In [8], an analysis of the normative framework for testing of MI SW at the national level was conducted to determine its suitability for carrying out conformity assessment. It has been established that DSTU 7363 [6] contains only general SW protection requirements and does not define the SW testing methodology.

National metrology institutes and conformity assessment bodies are interested in the existence of effective testing methodology for MI SW, risk assessment and application-related threats. Actual issues are the development of national SW testing procedures for MI SW taking into account all essential requirements of OIML document [1] and WELMEC guidelines [2-3]. At the same time, it is important to conduct an analysis of the practical testing of national SW testing procedures for MI SW with purpose of its suitability for use [9].

2. The Main Content of the Research

On the basis of the conducted analysis, a special SW testing methodology of MI SW was developed taking into account all essential requirements of OIML document [1] and WELMEC guidelines [2-3]. This methodology contains the methodology for testing of MI SW on the following generalized quality indicators:

K_P is indicator of software characteristics with built-in computer;

K_U is indicator of software characteristics with a universal computer;

K_L is indicator of checking storage devices;

K_T is indicator of verification of data transmission devices;

K_S is indicator of checking of reading;

K_D is indicator of checking of SW sharing levels;

K_I is a special indicator of SW testing for a specific MI.

For each of these generalized quality indicators (main requirements), the relevant partial indicators (specific requirements) are determined and the expressions are set for the numerical value of the indicator.

The numerical value for each generalized quality indicator is determined by the expression:

$$K_k = \sum_{i=1}^{M} K_i w_i^K / M, \quad (1)$$

where M is the total number of estimated partial indicators for a certain generalized indicator; K_i is the partial indicators for a certain generalized indicator; w_i^K is the weighting coefficient of the corresponding partial indicator.

In [10] a comparative analysis of the results of conformity assessment of MI SW was carried out. The generalized and particular quality indicators were selected for assessing the quality of MI software and expressions to obtain the numerical value of each partial indicator for each generalized indicator were developed. For comparative evaluation, the Analytic Hierarchy Process (AHP) was used, since it allows comparing and quantifying alternative solutions [11-12]. For pairwise comparison of all quantitative and qualitative indicators with the presentation of the equation in a quantitative form, the Saati scale was used.

In accordance with requirements of the WELMEC guidelines for testing MI SW at the beginning of the evaluation, it is necessary to determine to which basic configuration this SW is related to: with a built-in computer P or with a universal computer U. In the further, it is necessary to use the full set of requirements (quality indicators) that relate to the corresponding basic configuration. The K_P and K_U quality indicators are important by default.

In the developed methodology, all of the above indicators, its partial indicators and its numerical scales are used as in [9]. K_P has 7 partial indicators; K_U – 9; K_L – 8; K_T – 8; K_S – 3; K_D – 4; K_I – 6. In general, the total indicator K_{tot} for MI SW can be estimated as a simple sum of all generalized indicators. The higher a numerical value for MI SW, the more preferable it is MI SW.

The developed methodology for the testing of six varieties of MI SW with built-in computer P (meters of electric and thermal energy, gas-dispensing columns, gas chromatograph) was used. The testing results in the evaluation of compliance of SW1–SW6 by methodology are shown in Table 1 and Fig. 1.

The analysis of the obtained results shows the advantage of MI SW in the following sequence: SW1 > SW5 > SW2 > SW6 > SW3 > SW4. At the same time, the K_S and K_D indicators a have the same values for all MI SW. It is determined that those indicators are practically inapplicable and can be neglected.

Table 1. Results of testing SW for generalized indicators.

	K_P	K_L	K_T	K_S	K_D	K_I	K_{tot}
SW1	9	8	5	1	1	7	31
SW5	9	8	4	1	1	7	30
SW2	9	7	3	1	1	7	28
SW6	9	7	4	1	1	6	28
SW3	9	6	4	1	1	6	27
SW4	9	8	5	1	1	3	27

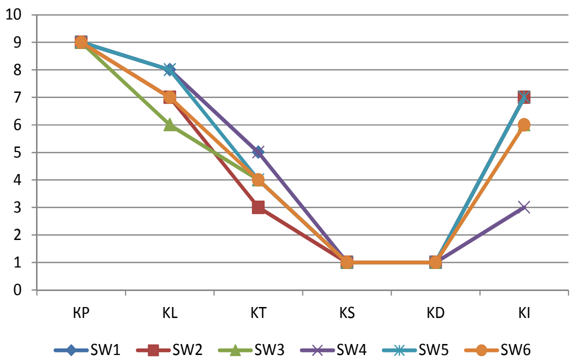

Fig. 1. Results of testing SW for generalized indicators.

In recent decades, Rasch mathematical model [13–17] is widely used to create new or to view existing scales. This model ensures that valid results are obtained through the use of statistics of adequacy, diagnostic information; presents test parameters on a single common linear scale, which helps in the criterion-oriented interpretation of the data.

The mathematical side and the very theory of G. Rasch have been successfully developed in [14]. If the data correspond to the Rasch model, then as a result, they are presented on an interval scale that is resistant to the loss of some primary data. Therefore, the model is a method of objective scaling of data.

Analysis of the data by the Rasch model is the matching test between the data received and the selected model. The measure of Rasch (the Rasch scale) is a linear value on the additive scale representing a hidden variable. In the Rasch model, the index of distribution of the subject, including the measurement error, is used. The magnitude of the measurement error is not uniform over the test range, but is usually greater for more extreme points (low and high). The Rasch scale of the successive response of each subject to each item has interval-scale properties [17].

Rush linear scales are initially expressed in units within 1, but can be redistributed according to normal scaling, from 0 to 100, while maintaining the aggregate addictiveness. The Rusch model also estimates the calculation errors at each level as standard measurement errors. The error is always greater at the upper and lower ends of the scale, since the Rush model is not bounded in boundaries, but measures from the middle of the range of values and assumes infinity in both directions. Measurements are better when the mean values of the elements are closer to the mean values of the scale, that is, the true estimate is more uncertain when the boundaries of the scale approach.

Special characteristic – logit is a key element for the probabilistic Rasch model [13]. This is the probability log unit – the unit of measurement used for calibrating the items and measuring the subjects by the hidden variable. Measurement means the location (usually in logits) by the hidden variable.

The logit of the number p is the probability determined by the formula:

$$\operatorname{logit}(p) = \log\left(\frac{p}{1-p}\right) = -\log\left(\frac{1}{p}-1\right) \quad (2)$$

The value of $p/(1 - p)$ is the corresponding coefficient and the logit of probability is the logarithm of the odds.

The logistic function of any number α is the inverse of the logit:

$$\operatorname{logit}^{-1}(\alpha) = \operatorname{logitic}(\alpha) = \frac{1}{1+\exp(-\alpha)} = \frac{\exp(\alpha)}{1+\exp(\alpha)} \quad (3)$$

Infit and Outfit statistics are the most widely used determinative statistics of the Rasch model. The Infit statistics are more critical when the scale of the item is close to the subject's scale, and the Outfit statistics are more critical if the indicators at the extreme limit of the scale are not metrics of the subject. Rasch charts and tables use normalized unweighted averages, so that the graphs are symmetric centered to zero [18].

The final stages of obtaining the characteristics of the Rusch model based on the best decisions are: uniform arrangement of the values of the elements (equality of steps of the scale); reduced measurement error (increased accuracy); the likelihood and improbability (suitability) of the elements and qualities of the subject expected of the model; overall reliability (noise – excessive data unpredictability, possibly due to excessive randomness or multidimensionality); simplicity; the nature of the elements being measured.

The Rasch model suggests that the probability of approving any category of responses to an object depends entirely on the subject's (SW) ability and complexity of the object.

3. Results of Comparative Analysis of Software Quality

The received primary data on these SW were processed using the software Winsteps Rasch 4.4.5, which implements the Rasch model. The results of the transformation of the input primary data by the items (indicators) and by the subjects (SW) in the Rasch measurement are shown in Table 2 and Table 3 respectively.

The results of the measurement by indicators and by SWs are presented in logits in descending order and are shown in Table 2 and Table 3. Measurement error is based on the Rasch model, that is, the estimated value, which, when added and subtracted from the measurement in logits, gives a minimum distance before the difference becomes significant.

Table 2. Results of data conversion concerning indicators.

Indicator	Measurement	Measurement error	Infit statistics		Outfit statistics	
			MNSQ	ZSTD	MNSQ	ZSTD
K_T	1.77	0.44	0.67	-0.43	0.91	0.07
K_I	0.07	0.39	1.13	0.42	0.93	0.05
K_L	-1.84	0.64	0.97	0.18	0.85	-0.03
K_P	Inestimable					
K_D	Inestimable					
K_S	Inestimable					
AV	0.00	0.49	0.93	0.10	0.90	0.00
SD	1.48	0.10	0.19	0.40	0.03	0.00

Table 3. Results of data conversion concerning SWs.

SW	Measurement	Measurement error	Infit statistics		Outfit statistics	
			MNSQ	ZSTD	MNSQ	ZSTD
SW1	1.46	0.68	0.09	-1.72	0.15	-1.35
SW5	1.02	0.66	0.36	-0.73	0.38	-0.70
SW2	0.18	0.65	0.91	0.18	0.77	-0.01
SW6	0.18	0.65	0.04	-1.96	0.06	-1.81
SW3	-0.24	0.64	0.72	-0.11	0.92	0.21
SW4	-0.24	0.64	3.38	2.12	3.08	1.83
AV	0.39	0.65	0.92	-0.40	0.90	-0.30
SD	0.64	0.01	1.15	1.40	1.02	1.20

The line AV is average value, the line SD is standard deviation. The columns of Infit and Outfit statistics contain parameters that characterize the matching of the data of the Rasch model: MNSQ is the value that characterizes the level of randomness of the results or the discrepancy between the measurement model data; ZSTD is standardized MNSQ values, that is, the probability of mean-square-statistics, expressed as z-statistics (mean-square deviation). MNSQ is also referred to as a relative xi-square or normalized xi-square.

Weighted average statistics of conformity are the xi-square statistics divided by its degrees of freedom. For the probability $p \leq 0.05$ (two-way distribution), ZSTD> |1.96|. The most expected values for MNSQ are near 1.0. The most qualitative and relevant values are MNSQ values ranging from 0.5 to 1.5.

Values below 1.0 indicate that the data are either too predictable, or excessively predictable, or overestimated model data. Values above 1.0 indicate too unpredictable data or underestimated model data. Values greater than 1.5 indicate uncertainty and "noise" (excessive unpredictability of data) in the input data, values less than 0.5 are also undesirable because they indicate an "information overload" of an item.

The MNSQ values from –2.0 to +2.0 are acceptable. The values of MNSQ for a module larger than 2.0 are considered to be non-conforming to the measurement model and cannot be used in the analysis of the results. The analysis begins with questions of high MNSQ value.

The obtained MNSQ values for the indicator for Infit statistics range from 0.67 to 1.13, and for Outfit statistics from 0.85 to 0.93 for K_T, K_I, and K_L indicators, but for K_P (all values are 9), K_D (all values are 1), and K_S (all values are 1) indicators are inestimable. This indicates that only K_T, K_I, and K_L indicators are acceptable for the analysis by the Rasch model. In this regard, it is considered necessary to clarify the feasibility of an expert assessment on K_P, K_D, and K_S indicators.

The obtained MNSQ values for the SW for Infit statistics range from 0.04 to 3.38, and for Outfit statistics from 0.06 to 3.08. This indicates that all these values are acceptable for the analysis by the Rasch model. Only for the SW4, the values of the Infit and Outfit statistics are respectively 3.38 and 3.08, which indicates the presence of "noise" in the input data.

The correlation coefficient (may take values from –1 to +1) is considered as a measure of reliability and validity, and is used to identify, refine, and possibly exclude poorly matched items. The standard deviation is the mean square root of the difference between a sample of values and a mean. The obtained correlation coefficient for indicators is equal from 0.25 to 0.62 (only for K_T, K_I, *and* K_L indicators) which indicates a medium correlation of the corresponding data. The obtained correlation coefficient for SWs is equal from 0.85 to 0.98 (for all SW) which indicates a very large correlation of the corresponding data.

Using the software Winsteps Rasch 4.4.5 [18], graphical reports were also obtained: characteristic curves, information functions, etc. In Fig. 2, characteristic curves of all evaluated SWs for all indicators, the analysis of the mutual placement of which helps to improve the evaluation as a system of indicators were constructed. In this case, most curves are concentrated in the middle and lower than average complexity. Characteristic curves practically uniformly fill the entire interval from –4.9 to +4.9 logits with the maximum allowable range from –5 to +5 logits. This indicates a fairly agreement between the indicators established for the evaluation of SW.

For each indicators and evaluation as a whole, you can get a graphical representation of the correspondence of the data of the selected model (Fig. 3). The obtained data indicate the presence of a correlation with the data for the selected model. Converted data for the evaluated SW according to the established indicators are shown in Fig. 3. This allows us to clearly show the ranking of SW based on the results of the application of the Rasch model for all established indicators.

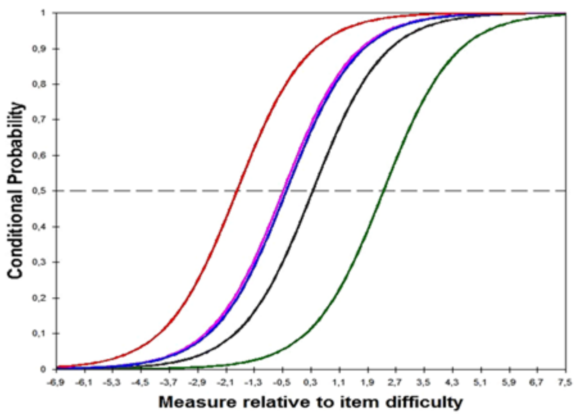

Fig. 2. Characteristic curves for evaluated SW.

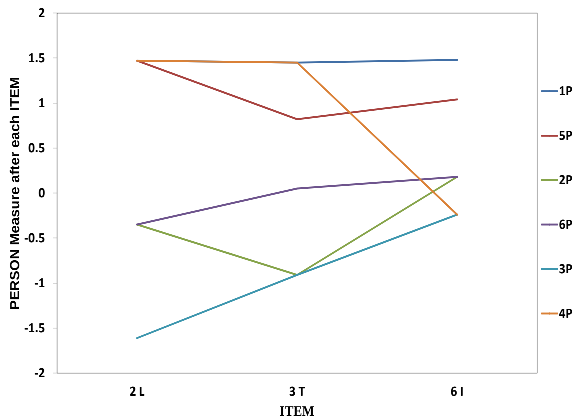

Fig. 3. Converted data on evaluated SW according to the indicators.

Differential Person Function (DPF) specifies the part of the item label which is to be used for classifying items in order to identify DPF – uniform or non-uniform – using the columns election rules. On Fig. 4 the DPF of all evaluated SWs for selected indicators, the analysis of the mutual placement of which helps to improve the evaluation as a system of indicators were constructed.

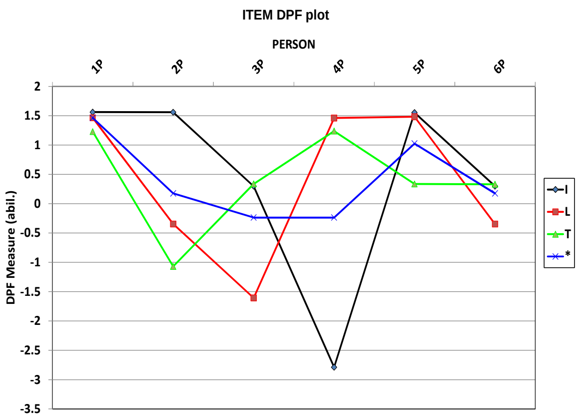

Fig. 4. DPF of evaluated SW according to the selected indicators.

On Fig. 5 a standard histogram of the person (indicator) ability distribution and item (SW) difficulty distribution displays. A Rasch-Thurstone threshold is a location on the latent variable with a precise definition. It is the point on the variable (in the context of a particular item) at which the probability of being observed below a given category is equal that of being observed in or above that category. Thurstone thresholds are independent of the number of categories in the rating scale.

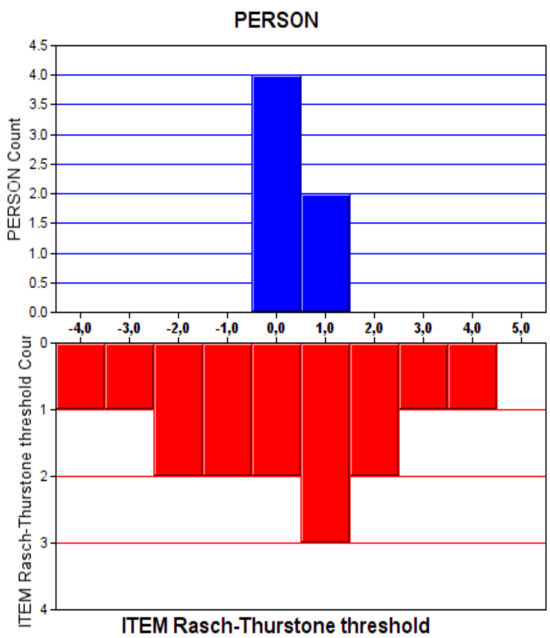

Fig. 5. Standard histogram of the indicator ability distribution and SW difficulty distribution.

4. Conclusions

The methodology that are suitable for evaluating the SW was considered in detail. The obtained results showed the possibility of applying the Rasch model for the analysis of the scale for evaluating the SW. A comparative analysis of the results obtained with the use of simple scores method and Rasch model showed convergence, suitability and correlation of the obtained values for SW.

References

[1]. OIML D 31:2008, General Requirements for Software Controlled Measuring Instruments, *OIML*, Paris, 2008, 53 p.

[2]. WELMEC 7.2. Software Guide (Measuring Instruments Directive 2004/22/EC). URL: http://www.welmec.org/fileadmin/user_files/publications/WG_07/Guide_7.2_2015_Software.pdf

[3]. WELMEC 2.3., Guide for Examining Software (Non-automatic Weighing Instruments). URL: http://www.welmec.org/fileadmin/user_files/publications/2-3.pdf

[4]. Velichko O. N., Basic tests, stages, and features in monitoring measuring instrument software, *Measurement Techniques*, Vol. 52, Issue 6, 2009, pp. 566–571.

[5]. Velichko O. N., Normative base for certification of measurement provision software, *Measurement Techniques*, Vol. 50, Issue 4, 2007, pp. 364–371.

[6]. DSTU 7363:2013, Prohramne zabezpechennia zasobiv vymiriuvalnoi tekhniky. Zahalni tekhnichni vymohy, *Minekonomrozvytku Ukrainy,* Kyiv, 2013, 11.

[7]. Velychko O., Gordiyenko T., Hrabovskyi O., Testing of measurement instrument software on the national level, *Eastern-European Journal of Enterprise Technologies*, Vol. 2, Issue 9 (92), 2018, pp. 13–20.

[8]. Velychko O., Gaman V., Gordiyenko T., Hrabovskyi O., Testing of measurement instrument software with the purpose of conformity assessment, *Eastern-European Journal of Enterprise Technologies*, Vol. 1, Issue 9 (97), 2019, pp. 19–26.

[9]. Velychko O., Hrabovskyi O., Gordiyenko T., Peculiarities of Testing of Measurement Instrument Software in Ukraine. Sensors and Electronics Instrumentation Advances: in *Proceedings of the 5th International Conference on Sensors Engineering and Electronics Instrumentation Advances (SEIA'2019)*, Adeje, Tenerife (Canary Islands), Spain, 25-27 September 2019, pp. 241–243.

[10]. Velychko O., Hrabovskyi O., Gordiyenko T., Quality assessment of measurement instrument software with analytic hierarchy process, *Eastern-European Journal of Enterprise Technologies*, Vol. 4, Issue 9 (100), 2019, pp. 35–42.

[11]. Saaty T. L., The Hierarchon: A Dictionary of Hierarchies, Analytic Hierarchy Process, Vol. 5 (Book 5), *RWS Publications*, Pittsburgh, Pennsylvania, 1992, 496.

[12]. Drake P. R., Using the Analytic Hierarchy Process in Engineering Education, *International Journal of Engineering Education*, Vol. 14, Issue 3, 1998, pp. 191–196.

[13]. Rasch G., Probabilistic models for some intelligence and attainment tests, Danish Institution for Educational Research, Copenhagen, 1960, 199 p.

[14]. Andrich D., Rasch models for measurement, *Sage Publications, Inc.,* Newbury Park CA, 1988, 95 p.

[15]. Wright B. D., Linacre J. M., A measurement is the quantification of a specifically defined comparison, Rasch model derived from objectivity, *Rasch Measurement Transactions*, Vol. 1, Issue 1, 1987, pp. 4–5.

[16]. Bond T. G., Fox C. M., Applying the Rasch model: fundamental measurement in the human sciences, 2nd ed., *Lawrence Erlbaum Associates*, Mahwah, New Jersey, 2007.

[17]. Velychko O., Gordiyenko T., A comparative analysis of the assessment results of the competence of technical experts by methods of Analytic Hierarchy Process and with using the Rasch model, *Eastern-European Journal of Enterprise Technologies*, Vol. 3, Issue 3 (93), 2018, pp. 14–21.

[18]. A User's Guide to WINSTEPS®MINISTEP Rasch-Model Computer Programs. Program Manual 4.4.0 by John M. Linacre, 2019, 704 p.