https://doi.org/10.21122/2227-1031-2024-23-4-345-354

УДК 53.087.2

Model for Ensuring the Reliability of Expert Quality Control of Products and Processes

P. S. Serenkov[1], V. M. Romanchack[1], E. A. Davidova[1], A. A. Hurynovich[1]

[1]Belarusian National Technical University (Minsk, Republic of Belarus)

Abstract. The reliability of the results of sensory analysis depends on a number of factors that affect the objectivity of the tests carried out. Today, the credibility of subjective measurements is primarily achieved through standardization. However, the issue of the credibility of subjective measurements remains, furthermore, it moves to a new level. Special attention must be paid to subjective measures related to the measurement of sensations to ensure credibility of results. The dynamics of increasing credibility through factor standardization lags behind the dynamics of stakeholder demand for increasing the credibility of subjective measurements. The purpose of the paper is to consider subjective measurements from the point of view of the development of the theory of quantitative measurements and to substantiate a process model for measurement that ensures the meaningfulness of the results in relation to expert assessments that ensure the subjectivity of measurements when conducting sensory tests, the results of which form decisions on compliance or non-compliance. The object of research is expert assessment methods used in sensory measurements, specifically in the evaluation of participating experts. The research methods used in this work include system analysis of measurement theories, method of alternatives, and standardized methods of expert assessment. A model of quantitative measurements is proposed to ensure meaningful measurement results, based on an analysis of the evolution of measurement theories. The problem of ensuring the meaningfulness of subjective measurements is formulated, which manifests itself in the form of risks of making incorrect decisions about characteristics of food products and processes based on expert assessments that lack reliability. An algorithm for quantitative measurements has been defined and tested on a specific example of expert assessment, demonstrating the importance of the identified problem of ensuring the reliability of expert assessments.

Keywords: sensory analysis, assessor, expert assessments, subjective measurement, meaningful measurement results, quantitative measurements theory, measurement scales

For citation: Serenkov P. S., Romanchack V. M., Davidova E. A., Hurynovich A. A. (2024) Model for Ensuring the Reliability of Expert Quality Control of Products and Processes. *Science and Technique*. 23 (4), 345–354. https://doi.org/10.21122/2227-1031-2024-23-4-345-354

Модель обеспечения достоверности экспертного контроля качества продукции и процессов

Докт. техн. наук, проф. П. С. Серенков[1], канд. физ.-мат. наук, доц. В. М. Романчак[1], канд. техн. наук Е. А. Давыдова[1], А. А. Гуринович[1]

[1]Белорусский национальный технический университет (Минск, Республика Беларусь)

Реферат. Достоверность результатов органолептического анализа зависит от ряда факторов, влияющих на объективность проводимых испытаний. Повышение достоверности субъективных измерений обеспечивается сегодня главным

Адрес для переписки
Серенков Павел Степанович
Белорусский национальный технический университет
просп. Независимости, 65,
220013, г. Минск, Республика Беларусь
Тел.: +375 17 331-11-20
pavelserenkov@bntu.by

Address for correspondence
Serenkov P. S.
Belarusian National Technical University
65, Nezavisimosty Ave.,
220013, Minsk, Republic of Belarus
Тел.: +375 17 331-11-20
pavelserenkov@bntu.by

образом за счет их стандартизации. Однако проблема достоверности субъективных измерений остается, мало того, переходит на новый уровень. Субъективные измерения, связанные с измерением ощущений, требуют особого внимания в контексте достоверности результатов. Динамика повышения достоверности за счет стандартизации факторов отстает от динамики спроса заинтересованных сторон на повышение достоверности субъективных измерений. Цель работы – рассмотреть субъективные измерения с точки зрения развития теории количественных измерений и обосновать модель процесса измерений, обеспечивающую осмысленность результатов в отношении экспертных оценок, обеспечивающих субъективность измерений при проведении органолептических испытаний, по результатам которых формируются решения о соответствии или несоответствии. Объектом исследований являются методы экспертного оценивания, используемые в органолептических измерениях и, в частности, при оценке экспертов, принимающих в них участие. В работе использованы методы исследований: системный анализ теорий измерений, метод альтернатив, стандартизованные методы оценки экспертов. По результатам анализа эволюции развития теорий измерений предложена модель количественных измерений, обеспечивающая осмысленность результатов измерений. Сформулирована проблема обеспечения осмысленности субъективных измерений, проявляющаяся в виде рисков принятия некорректных решений в отношении характеристик пищевой продукции и процессов по результатам экспертного оценивания в силу их недостаточной достоверности. Определен алгоритм количественных измерений, апробированный на конкретном примере экспертного оценивания, демонстрирующий значимость установленной проблемы обеспечения достоверности экспертных оценок.

Ключевые слова: органолептический анализ, испытатель, экспертные оценки, субъективные измерения, осмысленность результатов измерений, теория количественных измерений, шкалы измерений

Для цитирования: Модель обеспечения достоверности экспертного контроля качества продукции и процессов / П. С. Серенков [и др.] // *Наука и техника.* 2024. Т. 23, № 4. С. 345–354. https://doi.org/10.21122/2227-1031-2024-23-4-345-354

Introduction

Currently, sensory analysis is widely used in the food industry to provide information on the chemical composition and a comprehensive assessment of product quality. Obviously instrumental methods of analysis alone are insufficient for a complete assessment of product quality. This is evidenced by the fact that the chemical composition of the products may be similar, but the sensory characteristics of these products will differ significantly. Therefore, comprehensive product control is usually based on a combination of instrumental and sensory methods. If we take into account the advantages of sensory methods for assessing product quality (availability, speed, cost-effectiveness, proximity to consumer assessment), then it is quite clear that in certain conditions these methods become of paramount importance.

The quality of food products can be assessed using technical measuring instruments or on the basis of the subjective opinion of a competent person, known as an assessor [1].

Sensory assessment may be made by three types of assessors: “sensory assessors”, “selected assessors” or “expert sensory assessors”.

Assessor can be “naive assessor” who do not have to meet any precise criterion of selection or training, or a person who have already taken part in some sensory tests (“initiated assessors”).

“Selected assessor” is an assessor who have been selected and trained for the particular sensory test.

According to GOST ISO 5492 [2] an expert sensory assessor is a sensory assessor with a demonstrated sensory sensitivity and with considerable training and experience in sensory testing, who is able to make consistent and repeatable sensory assessments of various products.

It is common knowledge that the reliability of the results of organoleptic analysis depends on the psychophysical state of the expert, his experience, proficiency in methods of sensory analysis, level of training, sensory abilities, test conditions, etc. [3].

Currently, it is generally accepted that measurement is always the process of experimentally obtaining one or more values of a quantity that can be reasonably assigned to the value. Objective measurement is associated with measurement by technical means, and objective methods are those in which the effects of personal opinion are minimized. Subjective or psychophysical measurements are associated with the measurement of sensations, where a person plays the role of a measuring instrument, and the subjective method is a method based on personal opinions [2].

Therefore, when measuring sensations, it is important to pay special attention to subjective measures in order to ensure the reliability of the results. If the reliability of objective measurements is supported by technical means such as standards, traceability, comparisons, repeatability, and reproducibility, then the reliability of results for subjective measurements of sensations becomes critical. The determination of criticality depends on the relationship between the reliability of subjective measurements and the risk of making incorrect decisions based on the results of monitoring and testing.

Obviously, the need for reliable subjective measurements is becoming increasingly important as the number of interested parties grows, due to the ever-increasing appearance of new materials, environment systems and substances that require sensory assessment and hedonic tests.

In the last few years there has been an increasing interest in the use of expert methods for the assessment of product quality [4–6], but the issue of the reliability of the measurement results is often overlooked.

It is important to note that strategies to improve the reliability of objective and subjective measures have different emphases. This difference is not only determined by the measurement infrastructure, as mentioned above, but also by the degree of meaningfulness of the results.

A review into current research in the field of sensory analysis has revealed that the main method of improving the reliability of subjective measurements is through standardizing the factors that affect the objectivity of the tests. These include: methods for selecting and training experts, conditions for conducting sensory analysis, implementation of control measures relating to the measurement process, methods for processing and evaluating expert information, validating results, etc. [3, 4].

However, the issue of the credibility of subjective measurements remains, furthermore, it moves to a new level. The dynamic of increasing credibility through factor standardization lags behind the dynamic of stakeholder demand for increasing the credibility of subjective measurements.

The purpose of this paper was to consider subjective measurements from the point of view of the development of the theory of quantitative measurements and to substantiate a process model for measurement that ensures the meaningfulness of the results in relation to expert assessments that ensure the subjectivity of measurements when conducting sensory tests, the results of which form decisions on compliance or non-compliance.

Materials and methods

This study focuses on expert assessment methods used in sensory measurements, in particular, in the evaluation of experts involved in such measurements. The following research methods were used in the work: system analysis of measurement theories, method of alternatives, standardized methods of expert evaluation [4].

In the field of psychophysical measurements, the intuitive and acceptable characteristics of the meaningfulness of measurement results is considered as the adequacy of the numerical form of their representation to the real characteristics of the measurement objects [5].

It is generally accepted that objective measurements work with scales that are “strong” in terms of information (interval, proportional, absolute) [6–8]. They are also known as metric. The adequacy of the numerical form of their representation to the real properties of the measured objects is generally beyond doubt for the measurement results presented in these scales.

Subjective measures mainly use scales considered 'weak' in terms of the information they provide, such as nominal and ordinal scales. There are doubts about the adequacy of the numerical representation of measurement results in these scales to the real properties of the objects being measured [6–9]. Of interest from this perspective is the requirement to present information in a form that allows for a high level of comprehension during perception and use, as well as its dimensional and functional information properties [10].

The lack of a systematic approach to the problem of results credibility is indicated by the different priorities for improving the processes of objective and subjective measurement. This is also evidenced by the wide variety of concepts, theories, methods, and scientific schools [6–8, 11–12]. The basics of measurement have been established in the early measurement theories of Helmholtz, O. Hölder, N. Campbell et al. [13–15]. In the development of theoretical foundations (alt: basics), especially from the perspective of subjective measurement, it is worth mentioning classical theory (G. Fechner et al. [6, 15, 16]), P. Bridgman's theory of operationism [17], representational theory (S. Stevens et al. [6–8, 18–19]).

Within a generalized model of a measurement system, it was S. Stevens who formulated the measurement meaningfulness concept. He believed that a measurement system is defined when its three elements are defined (Fig. 1):

– an empirical system that includes physical objects, sensations, judgments, and the relationships between them, specified axiomatically;

– a numerical system in which logical-mathematical relationships are specified axiomatically;

– function *f*, which is a homomorphic mapping of an empirical system into a numerical system. This function enables to assess the relationships between tangible objects by analyzing the relationships between their numerical representations.

Essentially, function *f* – a set of rules that guarantees the accuracy of relationships within both empirical and numerical systems.

According to S. Stevens, it is important to have strict and agreed-upon rules for assigning numbers to objects for each type of measurement to ensure confidence in the measurements [16, 19]. J. Pfanzagl developed and generalized the representative theory of measurements [6, 20].

The theory of measuring physical quantities is currently undergoing a shift in emphasis towards the processing and transformation of measurement results. This can be seen in the work of A. Kolmogorov and other scientists who have made significant progress in the field of the conversion of measurement information [21, 22].

Consequently, we can distinguish two main directions in the development of measurement theory, each claiming independence. The measurement process has a generalized model that consists of two conditionally independent models (Fig. 2):

– model of empirical measurements of quantities,

– model for converting measurement information.

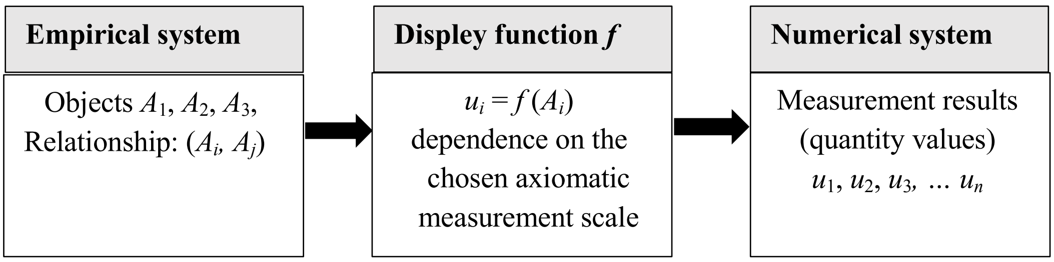

Fig. 1. Model of quantitative measurements as a generalized process of measuring a quantity

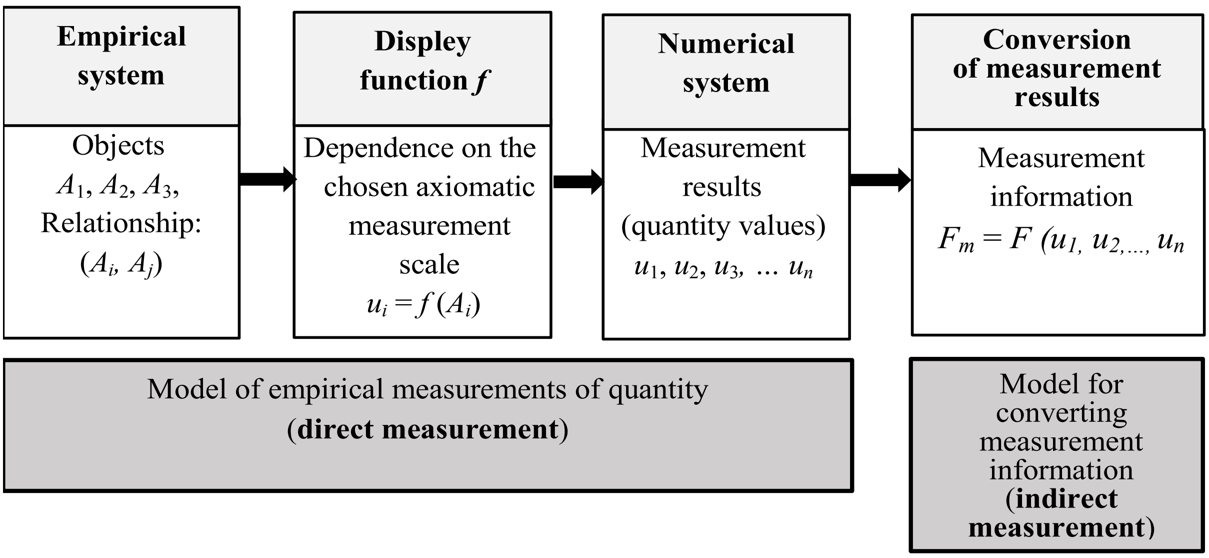

Fig. 2. Generalized model of the measurement process, including two conditionally independent models

The model of empirical measurement of a quantity is related to the model of direct measurement, in which the desired value of a quantity is obtained directly from the measuring instrument (in the case of subjective measurements – an expert). The measurement information transformation model is related to the indirect measurement model: the desired value of a quantity is determined from the results of direct measurements of other quantities functionally related to the desired quantity.

The development of the empirical measurement model faded into the background with the development of measurement information transformation models, which led to a number of issues concerning the credibility and meaningfulness of the results [9–10, 14, 22, 23].

Results and discussion

An analysis of existing measurement theories reveals that they are founded on a generalized model of the process of measuring a quantity (Fig. 1). Furthermore, the theories consider either the process model as a whole or its individual elements.

A structured analysis of the most common measurement theories, in accordance with the logic of Fig. 1, enables the identification of their shared weaknesses.

1. Measurement theories do not define the actual measurement procedure in an empirical system.

2. Measurement theories do not define a strict (natural) link between the empirical and numerical systems, which would enable us to claim that the measurement results are entirely meaningful. The measurement scale serves as the carrier of this connection.

3. Measurement theories do not fully solve the problem of ensuring the meaningfulness of measurements: they lack criteria for determining the adequacy of measured results in an empirical system compared to measured results in a numerical system (measurement scale).

Barzilai's theory of quantitative measurements is based on the relationships between objects, rather than their characteristics [15, 16]. And measurement is the process of assigning numerical values to relationships between objects, rather than to the objects themselves. Only in this case, according to J. Michel, real numbers are not assigned but naturally generated in the measurement process [24].

These principles are used to determine the correct measurement procedure. To achieve this, we will define an "exclusive group of objects" axiomatically, as proposed by A. Friedman [25], which will enable us to make a specific evaluation. Objects $A_1, A_2, \ldots, A_n$ are arranged in ascending order based on their measured values, which change uniformly in magnitude. This means that the empirical comparison of successive pairs of objects produces identical results. Then (we?) assign the value of the quantity u_i to each object A_i. It is assumed that for such objects the successive differences in value are equal to each other: $u_2 - u_1 = u_3 - u_2 = \ldots = u_n - u_{n-1}$.

In this case, the equality is true

$$u_i - u_j = \lambda_1 (i - j), \quad (1)$$

where $\lambda_1 > 0$, λ_1 – unknown constant.

Thus, during the process of measurement, real numbers, which represent the values of the quantity, are naturally obtained.

This kind of special assessment, following A. Friedman, we will call it measurement [25]. Thus, a mapping is defined that corresponds naturally to the empirical result of comparing a pair of objects using a numerical value – the difference in values. The measured quantity values are defined with the accuracy of a linear transformation, i.e. in the interval scale.

Let's assign a value v_i to each object A_i, and assume that successive relations of values are equal $v_2 / v_1 = v_3 / v_2 = \ldots = v_n / v_{n-1}$.

The result is:

$$\ln (v_i / v_j) = \ln(v_i) - \ln(v_j) = \lambda_2 (i - j), \quad (2)$$

where $\lambda_2 > 0$, λ_2 – unknown constant.

Thus, the second method of measurement is defined as mapping the results of an empirical comparison onto a set of results of an algebraic operation.

This mapping naturally matches the empirical result of comparing a pair of objects with a number – the ratio of values or the difference in logarithms of values. The measured quantity values are defined on a scale of logarithmic intervals [16, 25].

From equations (1) and (2), it follows that the values obtained on the interval scale and the logarithmic interval scale are related by the formula

$$(u_i - u_j) = \lambda \ln\left(v_i / v_j\right), \quad (3)$$

where $i, j = 1, 2, \ldots, n$; u_i and v_i – quantity values, $\lambda = \lambda_2 / \lambda_1$.

To avoid considering two measurement methods, it is convenient to introduce the concept of rating based on equality (3). Let's denote the left and right sides of equality (3) by the symbol R_{ij} and define two mappings or two measurement models:

$$R_{ij} = \lambda_1 (u_i - u_j), \quad (4)$$

$$R_{ij} = \lambda_2 \ln(v_i / v_j), \quad (5)$$

where R_{ij} – rating values, $i, j = 1, 2, \ldots, n$; u_i, v_i – quantity values obtained through various empirical measurement methods.

For objects with uniformly changing magnitude values, the rating is determined accurate to a scale constant λ using the formula:

$$R_{ij} = \lambda (i - j), \quad (6)$$

The classical definition of the rating follows from the adjusted model of S. Stevens [19, 20]. The rating is the result of measurements of the relations of the objects of the empirical system. The scale of these measurements can be determined using the rating.

J. Barzilai noted that the lack of agreement on the preference for particular measurement theories is mainly due to scaling errors. Scaling errors turn measurement into an operation that produces meaningless numbers [13, 14].

Identifying measuring scales was already done by S. Stevens in 1946 [19, 20]. However, his concept of scaling contains internal contradictions. This is due to the fact that the correspondence between the empirical and numerical systems (Fig. 1, 2) was determined intuitively without proper justification.

In our opinion, there are two points that prove the correctness of the scaling:

1) the measurement scale should be a natural consequence of the measurement procedure;

2) the empirical and numerical systems of the measurement model must be connected by isomorphism (Fig. 1).

Isomorphism is a mapping of systems that is mutually unambiguous. This means that the empirical system is equivalent to the numerical system. Therefore, the numerical system can be defined as a natural consequence of the empirical system, rather than axiomatically.

Every empirical measurement involves a comparison operation, which produces the result of an algebraic operation, such as the difference or ratio of values. The values themselves are naturally determined on an interval scale if they solve the system of equations (4), and on a logarithmic interval scale if they solve a system of equations (5). Additionaly, a ratio scale can be defined as an interval scale that includes a zero element, known as the origin.

The concept of a correct model of quantitative measurements is formed by the strict definition within the empirical system of the measurement procedure as a comparison operation and the natural consequent definition of the scale as the basis of the numerical system. This concept can be considered from unified positions for both subjective and objective measurements (Fig. 3). Here $A_1, A_2, \ldots, A_n, \ldots$ – the objects of measurement, u_i, v_i – respective numerical values of the objects. (A_i, A_j) – the outcome of empirical measurement of the relationships between these objects. The outcome of an empirical measurement is either the difference in values $(u_i - u_j)$, or the ratio of values v_i / v_j, which are transformed into a rating R_{ij}, the value of which does not depend on the measurement method. Based on the rating, the final measurement result U_{ij} is generated in interval scale or logarithmic interval scale, respectively.

The validity of the model is determined, among other things, by the fact that from expressions (4) and (5) the experimental laws of psychophysics by G. Fechner and S. Stevens can theoretically be obtained in the form of paired comparisons, and their equivalence can also be proven [25].

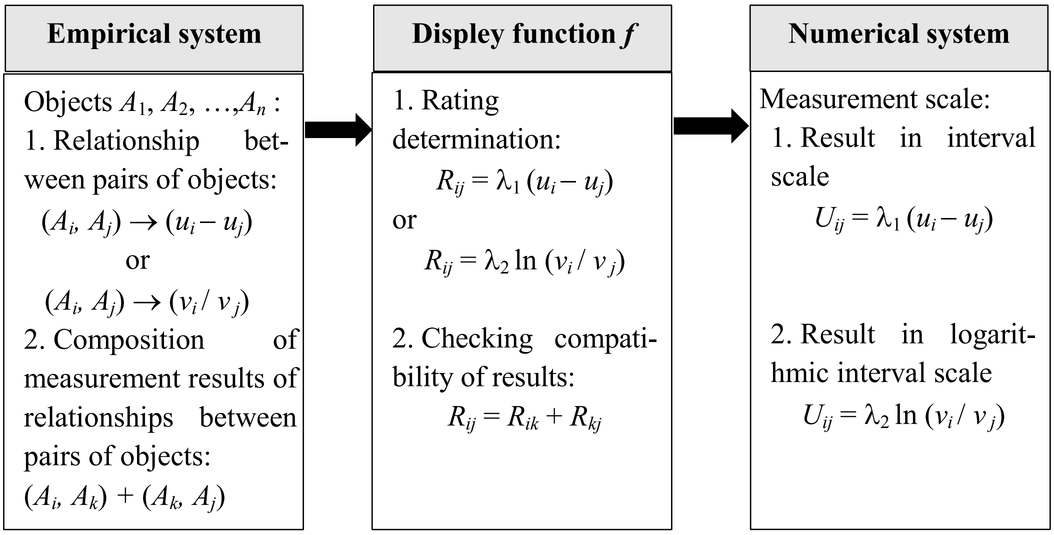

Fig. 3. Concept of a correct model of quantitative measurements

Fig. 3 shows expression $(A_i, A_k) + (A_k, A_j)$, which represents the composition of the results of empirical measurements of relations between pairs of objects. The results indicate the presence of the property of compatibility of results in the measurement model, as shown by the ratings R_{ik} and R_{kj}. The paper [25] demonstrates that rating values meet compatibility conditions in the form of:

$$R_{ij} = R_{ik} + R_{kj}, \tag{7}$$

The compatibility condition (7) in the numerical system can be considered as a criterion of adequacy of the measurement model, i.e. of the measurement results in the empirical system to the measurement results.

In practice, verifying all the compatibility equations (7) for systems of equations (4) or (5) can be a time-consuming task. The method of alternatives [16, 27] can be used to perform a partial verification of the compatibility equations.

For partial models (4) and (5), we have formulated the measurement algorithm as follows:

1. Select the measurement model (4) or (5);
2. Register the results of measurement $(u_i - u_j)$ or (v_i / v_j);
3. Calculate the ratings *Rij*;
4. Check the compatibility condition (7);
5. Depending on the measurement equation (4) or (5), the values of the measured quantity U_{ij} can be found (Fig. 3).

Let's use an example to demonstrate the proposed quantitative measurement model.

In order to implement the procedure of odour determination in a closed room, the suitability of potential experts and their olfactory ability should be checked in accordance to GOST ISO 16000-30 [27]. The standard regulates olfactory methods and criteria for assessing the ability to identify and distinguish odours from each other, as well as the threshold of odour perception. Appendix B of the standard provides a training methodology for confirming the olfactory ability of a certified expert.

Various methods are used to test a potential expert's olfactory skills. The program for analyzing odour intensity is based on a scale of intensity categories, which is implementted using an olfactometer according to GOST ISO 16000-30 [27].

To "calibrate" the trained sensory assessor, they must be presented with the smell of each intensity at least once. During subsequent analysis, each intensity is presented to tested experts at least twice in random order. The expert should assign each concentration of *n*-butanol to its correspondding intensity value.

The Q_ value is used as a criterion to assess the compliance of the certified expert with the requirements. It is calculated using the following formula:

$$Q\_value = \sum_{k=1}^{K} \frac{\sum_{i=1}^{j} (x_{ik} - I_k)^2}{j},$$

where x – individual result of a member of the commission; j – number of circles (one circle includes the assessment of all concentrations according to table 1); *k* – number of different concentrations according to table 3; *I* – intensity level according to GOST ISO 16000-30 [27]. The maximum *Q _ value* for a successful assessment should not exceed 6.

According to the given methodology, an expert certification was carried out, the results of which are shown in the table 1.

Table 1

Expert scores based on the GOST ISO 16000-30 model

The intensity of the odour of *n*-butanol, measured with an olfactometer according to GOST ISO 16000-30	Expert scores in points		
	Round 1 scores	Round 2 scores	Round 3 scores
0 – no odour detected	0	1	1
1 – very weak	3	2	2
2 – weak	3	3	4
3 – distinct	4	3	4
4 – strong	5	3	5
5 – very strong	5	5	5

The calculated value of the compliance criterion, *Q_value* = 6, 33, indicates that the expert did not pass the test.

How reliable is an expert's olfactory ability in producing legally significant results? Upon analysis of the empirical system of the measurement model proposed by GOST ISO 16000-30 [27], as shown in Fig. 3, it can be concluded that it is not clearly expressed. The system's objects, which consist of concentrations of odours from six categories, are clearly defined. Additionally, the relationships between these objects have also been defined. However, during empirical measurements of the expert's olfactory abilities, only identification (recognition) of the objects of the system "level of n-butanol odour intensity – score" is performed. It is known that the scores obtained can also be represented on a nominal scale.

The empirical system does not measure relationships between objects. Even if we assume that the expert has created a sequence of odor intensity levels during "calibration", we can only consider the same assessments on a rank scale. However, it is important to note that nominal and rank scales are not metric scales, and therefore estimates cannot be subjected to mathematical operations. Therefore, it could be argued that the *Q_value* criterion, which measures information according to Fig. 2, is not meaningful. This means that the assessment of conformity by the certified expert may not be reliable.

To ensure reliable conformity assessment, it is necessary to modify the empirical system of the measurement model. This can be achieved by adding a procedure for measuring the relations between objects, such as "by how much is one object superior to another?" or "by how many times is one object superior to another?".

To achieve these goals, in addition to the standard methodology (3 rounds of assessment), it is suggested to conduct three more rounds of empirical measurements following the algorithm outlined in the aforementioned article.

1. *Select the measurement model (5).* In each round of testing, the expert is presented with two odour samples of different intensity within one measurement. The expert should answer the question "by how many times is one odour sample more intense than the other?", i.e. it is proposed to apply the measurement model of S. Stevens [19, 20].

Note. In this case, the S. Stevens model was chosen because, in accordance with GOST ISO 16000-30 (Annex B), a number of intensity levels are built by geometric progression, i.e. so that any two adjacent levels differ by a factor of two.

Within each round of tests, comparative measurements of the odour intensity levels of *n*-butanol were arranged according to the plan "each of six concentration levels with one base level (any)".

Note. Each of the three cycles of comparative measurements can be carried out according to the same plan or according to different plans, e. g. "each of the six concentration levels with the previous intensity level (second with the first, third with the second, etc.)".

2. *Record the results of ratio measurements* (v_i / v_j). We choose the base level j = 3 and take the value of the *n*-butanol concentration at base level v_j = 1 (table 2).

Note. Row evaluations with a zero level are not filled (relational operations with zero are meaningless). A score of 4, for example, in table 2 means that the expert decided that the fifth level is 4 times larger than the third level. A score of 1/3 in table 2 means that the expert decided that the second level is 3 times smaller than the first.

Table 2

Expert scores of the S. Stevens model

The intensity of the odour of *n*-butanol, measured with an olfactometer according to GOST ISO 16000-30 model	Results of expert's measurements of (v_i/v_j) relation		
	Round 1 scores	Round 2 scores	Round 3 scores
0 – no odour detected	–	–	–
1 – very weak	1/4	1/9	¼
2 – weak	1/2	1/3	¼
3 – distinct	1	1	1
4 – strong	4	3	2
5 – very strong	7	5	6

3. *We calculate the ratings* R_{i3} according to (5) using the adjusted formula:

$$R_{i1} = ln\ (v_i\ /\ v_1\ )\ /\ ln(2) + 1, \quad (8)$$

where v_i – unknown values of the sample concentration level, i = 1–5.

According to the logic of the proposed measurement model (Fig. 3), on the basis of the rating values we form the final measurement results U_i on the scale of logarithmic intervals, the scale of which corresponds to the scale of point categories and form table 3.

Table 3

Results of measurements of the odour intensity level of *n*-butanol on the scale of logarithmic intervals

Odour intensity level of *n*-butanol odour measured by olfactometer according to GOST ISO 16000-30 model	Final measurement results U_i		
	Round 1 results	Round 2 results	Round 3 results
0 – no odour detected	0,0	1,0	1,0
1 – very weak	1,0	1,0	1,0
2 – weak	2,0	2,6	1,0
3 – distinct	3,0	4,2	3,0
4 – strong	5,0	5,8	4,0
5 – very strong	5,8	6,5	5,6

Note. Estimates for the zero level row are taken from table 1.

4. *We perform a partial test of the jointness equations* using the method of alternatives [14, 26]. The criterion for accepting the hypothesis about the consistency of the expert's estimates for each round of tests obtained by different methods was the condition of statistical significance of the correlation coefficients between the estimates according to GOST ISO 16000-30 measurement model [27] and the results of measurements according to the S. Stevens model. The correlation coefficients for all three rounds of testing are significant by Student's criterion at the 0.05 significance level, so the hypothesis of consistency of the experts' estimates is accepted.

This fact allows us to calculate the value of the expert compliance criterion *Q_value* = 4,0 based on alternative measurement results. The criterion indicates that the expert has been tested, i.e. meets the requirements.

The conflict arising from subjective measurements cannot be experimentally verified. The only way to verify is through theoretical justification of one or another measurement model. In this example, the authors appear to favour the alternative measurement model as it aligns with the general provisions of classical and modern measurement theories in terms of ensuring result awareness (Fig. 3).

CONCLUSIONS

This text discusses the issue of meaningfulness in measurements, specifically the subjectivity of measurements. It proposes a model of quantitative measurements based on an analysis of the evolution of measurement theories, which ensures the meaningfulness of measurement results. The model is based on two measurement methods. A special parameter, the rating R_{ij}, is associated with the difference or ratio of the sought values of the quantities of at least a pair of objects u_i and u_j, and is empirically measured within these methods. The assumption that both measurement models can be used together to measure the same quantity is justified. And the measurement results will be equivalent in a certain sense. An algorithm for quantitative measurements is formulated, as well as a reflection principle that ensures compliance between the empirical and numerical systems of the model.

The problem of ensuring the meaningfulness of subjective measurements is formulated, which manifests itself in the form of risks of making incorrect decisions regarding the characteristics of food products and processes based on the results of expert assessment due to their lack of reliability.

The analysis of the evolution of measurement theories has revealed a hidden component of the loss of reliability in subjective measurements.

This is due to the fact that existing measurement methods, including standardized ones, do not define the measurement procedure as a comparison operation. An evidence-based quantitative measurement model is proposed. The model ensures the meaningfulness of the results by measuring a special parameter – rating R_{ij}, which associated with the difference or ratio of the desired values of at least a pair of objects u_i and u_j. And the measurement results will be equivalent in a certain sense.

The concept of a correct quantitative measurement model is formed by the strict definition within the framework of the quantitative measurement model of the measurement procedure as a comparison operation and the natural definition of the scale that follows from it as the basis of a numerical system.

A quantitative measurement algorithm has been developed and tested using expert assessment as an example, demonstrating the importance of ensuring the reliability of expert judgement.

REFERENCES

1. State Standard ISO 6658-2016. *Sensory analysis – Methodology – General guidance, IDT.* Moscow, Standartinform Publ., 2016. 20 (in Russian).
2. State Standard ISO 5492–2014. *Sensory Analysis. Vocabulary*. Moscow, Standartinform Publ., 2015. 50 (in Russian).
3. Lovkis Z. V., Morgunova H. M., Shevchenko V. I., Davydova H. A. (2018) Sensory Analysis of Quality of Food Products. Requirements to Assessors. *Food Industry: Science and Technology*, 11 (1), 13–19 (in Russian).
4. State Standard ISO 8586–2015. *Sensory Analysis. General Guidelines for the Selection, Training and Monitoring of Selected Assessors and Expert Sensory Assessors*. Moscow, Standartinform Publ., 2015. 25 (in Russian).
5. Hubbard D. W. (2010) *How to Measure Anything: Finding the Value of Intangibles in Business*. 2rd ed. New York, John Wiley & Sons. 432. https://doi.org/10.1002/9781118983836.
6. Pfanzagl J. (1971) *Theory of Measurement*. New York, John Willey and Sons. 968. https://doi.org/10.1007/978-3-662-41488-0.
7. Luce R. D., Krumhansl C. L. (1988) Measurement, Scaling, and Psychophysics. Atkinson R. C. (ed.) [et al.]. *Stevens' Handbook of Experimental Psychology.* N.Y., John Wiley & Sons, 3–74.
8. Knorring V. G. (1980) Development of Representative Measurement Theory. *Izmereniya, Kontrol', Avtomatizatsiya: Nauchno-Tekhnicheskii Sbornik Obzorov* [Measurement, Control, Automatization: a Collection of Scientific and Technical Reviews]. Moscow, All-Union Scientific Research Institute of Information and Economics (INFORMPRIBOR), (11–12), 3–10 (in Russian).
9. Tolstova Y. N. (2018) Correlation of Theoretical and Empirical Knowledge when Using Mathematical Methods in Sociological Research. *Sociological Studies*, (12), 39–48. https://doi.org/10.31857/S013216250003164-3 (in Russian).
10. Gusev A. N, Utochkin I. S. (2011) *Psychological Measurements. Theory. Method*. Moscow, AspektPress Publ. 317 (in Russian).
11. Romanchak V. M. (2018) The Subjective Measurement of Probability. *Informatics*, 15 (2), 74–82 (in Russian).
12. State Standard 43.0.4–2009. *Informational Ensuring of Equipment and Operational Activity. Information in the Technical Activity. General Principles*. Moscow, Standartinform Publ., 2018. 20 (in Russian).
13. Corbetta P. (2003) *Social Research: Theory, Methods and Techniques*. New York, Sage. 336. https://doi.org/10.4135/9781849209922.
14. Barzilai J. (2005) Measurement and Preference Function Modelling. *Transactions in Operational Research*, 12 (2), 173–183. https://doi.org/10.1111/j.1475-3995.2005.00496.x.
15. Barzilai J. (2010) Preference Function Modelling: The Mathematical Foundations of Decision Theory. Ehrgott M., Figueira J.R., Greco S. (eds). *Trends in Multiple Criteria Decision Analysis*. New York, Springer, 57–86. https://doi.org/10.1007/978-1-4419-5904-1_3.
16. Krantz D. H., Luce R. D., Suppes P, Tversky A. (1971) *Foundation of Measurement. Vol. 1*. New York, Academic Press. 606.
17. Staddon J. E. (1978) Theory of Behavioral Power Functions. *Psychological Review*, 85, 305–320. https://doi.org/10.1037/0033-295x.85.4.305.
18. Bridgman P. W. (1951) *The Logic of Modern Physics*. New York, Macmillan Company. 228.
19. Stevens S. (1946) On the theory of Scales Measurement. *Science*, 103 (2684), 677–680. https://doi.org/10.1126/science. 103.2684.677.
20. Stevens S. (1957) On the Psychophysical Law. *Psychological Review,* 64 (3), 153–181. https://doi.org/10.1037/h0046162.
21. Kolmogorov A. N. (2006) Contemporary Debates on the Nature of Mathematics. *Problems of Informational Transmission*, 42, 379–389. https://doi.org/10.1134/S0032946006040107.
22. Levin S. F. (2012) Philosofical Problems and Statistical Methods of Fundamental Methrology. *Metafizika = Metaphysics,* (3), 89–118 (in Russian).
23. Shishkin I. F. (2016) Measurements of Non-Physical Quantities. *Journal of Physics: Conference Series,* 772, 1–7. https://doi.org/10.1088/1742-6596/772/1/012029.
24. Michell J. (1997) Quantitative Science and the Definition of Measurement in Psychology. *British Journal of Psychology*, 88 (3), 355–383. https://doi.org/10.1111/j.2044-8295.1997.tb02641.x.
25. Fridman A. A. (1965) *The World is Like Space and Time*. Moscow, Nauka Publ. 110 (in Russian).
26. Romanchak V. M. (2017) Model of Rating of Non Physical Quantity. *System Analysis and Applied Information Science,* (4), 34–44 (in Russian). https://doi.org/10.21122/2309-4923-2017-4-39-44.
27. State Standard ISO 16000-30–2017. *Indoor air. Part 30. Sensory testing of indoor air*. Moscow, Standartinform Publ., 2018. 30 (in Russian).

Received: 03.04.2024
Accepted: 09.06.2024
Published online: 31.07.2024