INTERNATIONAL STANDARD

ISO 11056

Second edition
2021-05

Sensory analysis — Methodology — Magnitude estimation method

Analyse sensorielle — Méthodologie — Méthode d'estimation de la grandeur

Reference number
ISO 11056:2021(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 12, *Sensory analysis*.

This second edition cancels and replaces the first edition (ISO 11056:1999), which has been technically revised. It also incorporates the amendments ISO 11056:1999/Amd.1:2013 and ISO 11056:1999/Amd.2:2015. The main changes compared with the previous edition concern the statistical treatment of the examples in Annex B:

— the Assessor factor is considered as fixed factor or as random factor (in the previous edition, the Assessor factor was always considered as a fixed factor);

— the R commands used to process the examples and to obtain the different tables are given explicitly (in the previous edition, only the tables of results were given);

— the numerical examples have been preserved without any modification to allow the user to understand the evolution in the processing of the tables;

— a new example has been added as B.2.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Magnitude estimation is a psychophysical scaling technique where assessors assign numerical values to the estimated magnitude of an attribute. The only constraint placed upon the assessor is that the values assigned should conform to a ratio principle, i.e. if the attribute appears to be twice as strong in sample B in comparison with sample A, the value assigned to sample B has to be twice that assigned to sample A. Attributes such as intensity, pleasantness or acceptability may be assessed using magnitude estimation.

Magnitude estimation method is often considered as being less susceptible to "end-effects" than the methods which employ an experimenter-defined continuous or discontinuous response scale. These "end-effects" occur when the assessors are unfamiliar with the extent of the sensations elicited by the products. Then assessors can assign one of the initial samples to a category which is too close to one of the ends of the scale. Consequently, they then find themselves short of graduations and are obliged to classify samples perceived as being different into the same category. This should not occur with magnitude estimation since, in theory, there are an infinite number of categories.

Allowing each assessor to start the process at any numerical value, i.e. to use their own scale, gives rise to a particularly important "assessor" effect. However, there are various ways of solving this problem:

— the analysis of variance (ANOVA) allows the "assessor" effect and the interactions to be taken into account;

— the assessors can be forced to a common scale by use of a reference sample to which a value has been assigned;

— the data supplied by each assessor can be reduced to a common scale by applying one rescaling methods.

It is up to the experimenter to choose the most appropriate approach based on the circumstances.

Magnitude estimation is the privileged method to determine the Steven's equation psychophysical power function. It can also be used to solve concrete problems.

NOTE The magnitude estimation method is not the most efficient technique for determining small differences between stimuli or for conducting assessments in the vicinity of a detection threshold.

EXAMPLE 1 A company produces a moderately successful beverage, but recent products which are sweeter, produced by a competitor, have made inroads into their shares of the market. It is decided to increase the sweetness level by one third in an attempt to recapture some of the market loss. In formulating the new product, knowing the power function of the sweetener will provide an estimation of the amount of sweetener necessary to reach the one third increase in sweetness level.

EXAMPLE 2 In the formulation of the new diet beverage, the intensity of the desired sweetness is known, but it is not yet decided whether to use aspartame or sucrose as a sweetener. Knowing the power functions of each substance, the iso sweetness lines can be plotted to determine the concentrations of each sweetener necessary for the desired sweetness level. This information coupled with cost/volume information can help inform the decision about which sweetener is more cost effective.

The calculations in Annex B were performed using R functions. Access to R packages is free. This information is given for the convenience of users of this document and does not constitute an endorsement or recommendation by ISO of the exclusive use of R packages. Other software may be used to perform the calculations required by this document.

The files are in the ME folder under the USB DISK H (format Text (separator: tabulation)).

The results can sometimes vary due to rounding errors, depending on the software used.

Sensory analysis — Methodology — Magnitude estimation method

1 Scope

This document specifies a method for applying magnitude estimation to the evaluation of sensory attributes. The methodology specified covers the training of assessors, and obtaining magnitude estimations as well as their statistical interpretation.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 3534-3, *Statistics — Vocabulary and symbols — Part 3: Design of experiments*

ISO 4121, *Sensory analysis — Guidelines for the use of quantitative response scales*

ISO 5492, *Sensory analysis — Vocabulary*

ISO 6658, *Sensory analysis — Methodology — General guidance*

ISO 8586, *Sensory analysis — General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors*

ISO 8589, *Sensory analysis — General guidance for the design of test rooms*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 3534-3, ISO 5492 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
magnitude estimation
process of assigning values to the intensity of a sensation elicited by a product attribute, or to its hedonic value, so the ratio between the assigned value and the assessor's perception of the attribute remains the same

3.2
external reference
first sample presented as a reference in a sample series in relation to which all subsequent samples are then assessed

3.3
internal reference
first sample introduced in a sample series in relation to which all subsequent samples are then assessed, and presented to the assessor as if it was a test sample

3.4
modulus
numerical value assigned to the *external reference* (3.2), which can be defined by the person conducting the test (fixed modulus) or left to the assessor to choose (non-fixed modulus)

3.5
rescaling
process that consists of multiplying the raw data supplied by each assessor by a factor which reduces the data of all the assessors to a common scale

Note 1 to entry: Adding the logarithm of this factor to the logarithm of the raw data is an equivalent process.

3.6
Steven's equation psychophysical power function
relationship that is expressed as:

$$R = KS^n$$

where

R is the assessor's response (e.g. perceived intensity);

K is the constant which reconciles the units of measurement used for R and S;

S is the stimulus (concentration of a chemical substance or physical variable);

n is the exponent of the power function and the slope of the regression curve for R and S when they are expressed in logarithmic units.

Note 1 to entry: In practice, Stevens's equation is generally transformed into natural logarithms:

$$\ln R = \ln K + n \ln S$$

4 Principle

Samples are presented successively to assessors, who are requested to record the intensity of an attribute of each sample by complying with the ratio principle.

The values are assigned by referring to the value of the first sample of the series. For this first sample, either each assessor is free to assign a value to it, or the value is fixed by the person conducting the test. The latter case is called "fixed modulus".

5 General test conditions

For the general test conditions, such as those concerning the facilities, preparation, presentation and coding of samples, International Standards on general methodology shall be followed, in particular ISO 6658 and ISO 8589, as well as those describing the methods using scales and categories, in particular ISO 4121.

6 Selection and training of assessors

6.1 General conditions for selection and training

The general conditions for selection and training shall be in accordance with ISO 8586.

As in all other sensory analysis methods, it is the responsibility of the panel leader to judge the required level of proficiency of the assessors. The objectives of the test, the availability of the assessors, the costs incurred by recruiting additional assessors, as well as their training, shall be taken into account when planning a training programme. Assessors are generally able to use the magnitude estimation methodology after three or four training tests.

6.2 Training specific to the magnitude estimation method

6.2.1 The assessment of surface areas of geometric shapes has been proved to be particularly suited for introducing assessors into the basic concepts of magnitude estimation. A set of 18 shapes (see Table 1) comprising six circles, six equilateral triangles and six squares ranging in size from approximately 2 cm^2 to 200 cm^2 has been used successfully for training assessors.

For the consumer panels, a shorter version may be used. For example, the training can be limited to area estimations.

Table 1 — Dimensions and areas of the training exercise shapes

Circles		Triangles		Squares	
Radius	Surface area	Side	Surface area	Side	Surface area
cm	cm^2	cm	cm^2	cm	cm^2
1,4	6,2	2,2	2,1	3,2	10,2
2,5	19,6	4,1	7,3	4,2	17,6
3,7	43,0	7,6	25,0	8,5	72,3
5,4	91,6	12,2	64,4	11,1[a]	123,2
6,8	145,3	15,5	104,0	11,1[a]	123,2
8,3	216,4	19,2	159,6	14,2	201,6

[a] Two 11,1 cm squares are introduced into the series in order to be able to evaluate the reproducibility of the assessors.

6.2.2 Prior to presenting the shapes to the assessors, instruct them in the principles of the method. This instruction shall include, but is not necessarily limited to, the following three points:

— the values shall be assigned on a ratio basis: if the attribute is twice as intense, a value twice as high shall be assigned to it;

— there is no upper limit to the scale;

— the value 0 shall be assigned only in the exceptional case where the attribute is not perceived.

Warn the assessors, at the time of training, that the general tendency is often to use round numbers (such as 5, 10, 20, 25, etc.) but that, with this method, all numbers are permitted and may be used.

As assessors are also influenced by the ratios mentioned during training, always take care to suggest to them the use of different ratios, e.g. 3/1, 1/3, 7/5, 5/6 without limiting oneself to 2/1 or 1/2.

6.2.3 Assign codes to the shapes and present the shapes separately by placing them in the centre of a sheet of white paper of approximately A4 size (21 cm × 29,7 cm).

Instruct each assessor to conduct the magnitude estimation, beginning the series with the presentation of the 8,5 cm square (external reference). Record the responses.

Depending on the procedure adopted for the test phase, train the assessors with or without a fixed modulus. With a fixed modulus, the person conducting the test assigns a value between 30 and 100 to the 8,5 cm square.

Without a fixed modulus, leave the assessors to assign the value of their choice to the first figure, but advise them not to choose too small a value.

Randomly present the geometric shapes prior to each test, so that their shapes and dimensions do not form a particular pattern.

6.2.4 After completing the assessment of the set of shapes, allow the assessors to compare their results with the average results of the group. If this is not practical, carry out this comparison with respect to the results obtained by a previous group.

The objective is to provide positive feedback to reassure the panellists that they understand the exercise. Care should be taken not to create the impression that there is a "right" answer. Unless their results are very different, departures from the group results should be explained as order effects; that is, their responses are affected by the order in which they evaluate the samples. They should be reassured that, despite individual order effects, the group's results will be accurate.

If the results of some assessors are very different, explain once again to these assessors the principles of the method.

6.2.5 When an assessor has successfully completed the area estimation exercise, further training should be given based on the product or type of substance that will be assessed in the actual test. This gives the assessor experience in applying magnitude estimation to attributes characterizing the test substance. The panel leader may need to design exercises for training panellists to identify correctly the attributes to be evaluated. This training may be drawn up using the general guidelines given in ISO 8586.

7 Number of assessors

7.1 General

As for the other methods employing scales, the required number of assessors depends on:

— how close together the various test products are in the attribute being evaluated;

— the training received by the assessors;

— the importance that will be attached to the decision following the test results (see ISO 8586);

— the objectives that can be identified in terms of statistical power.

In the absence of such distinctly identifiable objectives, refer to the recommendations given in 7.2 and 7.3.

7.2 Analytical and research panels

The panels shall be made up as given in Table 2.

Issues of statistical power need to be resolved on the variance of individual evaluations and the magnitude of the differences which need to be detected.

Table 2 — Formation of panels

Types of assessors	Minimum number of assessors	Recommended number
Experienced assessors, highly trained in the product and in the assessment of the attribute being studied	5	10
Experienced assessors, trained in the product and in the assessment of the attribute being studied	15	20 to 25
Newly trained assessors	20	20 and over

7.3 Consumer panels

The magnitude estimation method can also be used with consumer panels or for conducting market research studies. The number of persons to be selected shall then be determined on the basis of the population sampling requirements connected with these types of tests. The use of the magnitude estimation method does not offer any particular advantage in terms of the number of assessors required, and this number shall be the same as for a typical consumer type test, namely at least 60 persons and often much more.

8 Procedure

8.1 Presentation of samples

All the samples shall be presented in an identical manner (i.e. identical serving vessels and the same quantity of product).

The vessels containing the samples shall be coded, preferably using randomly selected three-digit numbers.

8.2 External reference sample

It is desirable that the reference sample has, for the attribute being studied, an intensity that is close to that of the geometric mean of all of the products under test.

NOTE A reference that can present an extreme value for the attribute would introduce a distortion.

One or more randomly coded reference samples may be included in the test series without the assessors being informed. This allows assessment of the repeatability of the assessor within the session.

8.3 Order of presentation of samples

The samples shall be presented all at once or in a sequential way to the assessors. The assessors shall follow the order indicated. As in all sensory tests, this order differs from one assessor to the other, the ideal situation being that the orders of the samples are balanced.

The panel leader can refer to tables proposed in Reference [3], which uses Latin squares to balance the design for order and carryover effects. Where this is not possible, use random order.

8.4 Magnitude estimations

8.4.1 General

Carry out the test in accordance with one of the techniques described in 8.4.2 to 8.4.4.

Questionnaire models for the reference sample are given in Annex A.

8.4.2 Without fixed modulus for the external reference

Each assessor evaluates the reference and assigns a value to it. Advise the assessors not to choose too small a value.

The assessor then evaluates each subsequent coded sample, comparing it with the reference, and assigns to it a value in relation to that which he/she has previously assigned to the reference.

8.4.3 With fixed modulus for the external reference

The panel leader specifies to the assessor that the reference sample has a value of, for example, 30, 50, 100, or whatever seems appropriate to the panel leader.

The leader instructs the assessor to make his or her subsequent judgements relative to the value assigned to the reference (fixed modulus).

8.4.4 Without external reference

It is possible to use magnitude estimation methodology without using any external reference sample. Due to the limits of the sensory systems (memory), it may be difficult for the assessors to refer systematically to the first sample. There are two possible cases, as follows.

a) The assessors are not forced to re-evaluate the first sample prior to evaluating each of the subsequent samples.

   It is then advisable to encourage the assessors to memorize the degree of the attribute being studied for this reference sample and to re-evaluate this reference if it appears necessary to them.

   It is therefore possible:

   1) prior to the test: to choose the presentation design in such a way that the first sample is not the same for each of the assessors; ideally, each sample should be used as a reference for an equal number of assessors; the variances of the mean differences between samples will therefore be equal;

   2) at the time of analysis: to apply an arbitrarily very high (theoretically infinite) weighting to the evaluations of the first sample of each assessor, so that the variances of the differences are correctly estimated.

b) The assessors are asked to evaluate each sample by comparing it to the immediately preceding sample.

   NOTE The problem that then arises is that, for each assessor, the evaluation errors are autocorrelated, and the variance of the difference of two successive samples will be smaller than the variance of the difference of two non-successive samples.

   It is therefore possible:

   1) prior to the test: to choose the presentation design in such a way that all the possible permutations of the samples are presented to an equal number of assessors; if this is not possible, try to propose orders which approximate best this ideal model; the variances of the mean differences between samples will then be equal or, at least, fairly close;

   2) at the time of analysis: to employ autocorrelated error models, the methodology of which is, however, slightly more complicated.

It is to be noted that even if one proceeds as proposed in case a) (systematic comparison with the first sample in order to carry out the evaluation), an autocorrelation term, linked to the evaluation of the preceding sample, however small it may be, very probably remains (this is also true, incidentally, for the tests with reference described in 8.4.2 and 8.4.3). The advice given earlier that the orders of the samples are balanced is therefore valid in all cases.

9 Analysis of data

9.1 Choice of data analysis method

The method of analysis depends on (see Annex B):

— the experimental design: complete design or incomplete design;

— the presence or not of replications;

— the status attributed to the Assessor factor (fixed factor or random factor), the Treatment factor being very generally considered as a fixed factor.

9.2 Presentation of raw results

The results may be presented in the form of a dual-entry table, placing horizontally the responses of the assessors after logarithmic transformation, and vertically the different samples.

When all the assessors have given a score the same number of times for each of the samples, a complete balanced plan is obtained and the model together with the assessor effect is orthogonal. If certain products have not been evaluated the same number of times by all the assessors, an incomplete plan is obtained and the model together with the assessor effect is non-orthogonal.

Since one cannot take the logarithm of zero, any zero response causes a problem. Different approaches have been used to deal with zeros. Zero values should be replaced by very small values. The specific value chosen should take into account the scale used by each panellist (e.g. half of the smallest value assigned by that panellist).

9.3 Establishment of product differences

An ANOVA, which explicitly accounts for all blocking factors (including unbalanced or nonorthogonal factors) and is carried out on logarithmically transformed data, is the most accurate method. In practice, it is not always possible to conduct an experiment leading to a complete design where all critical factors are balanced and orthogonal. For example, when a project extends over multiple sessions, it may be not possible to assemble exactly the same group of panellists at each session. It is always advisable to ask a statistician to set up the best possible experimental design.

When significant differences between products are revealed by ANOVA, one of the usual tests of multiple comparison of means is then carried out. An example of comparison of products without rescaled data on a complete design is given in B.1.

9.4 Regression

In cases where the values of a related variable S (such as concentration, physical quantity) are known as being capable of relating to the response R, it is possible to assume that Stevens's law is followed and to estimate its parameters by carrying out the linear regression of the sensory responses regarding this physical or chemical variable, according to Formula (1):

$$\ln R = \ln K + n \ln S \tag{1}$$

In such an analysis, the parameter that is of greatest interest is the slope that corresponds to the value n in Stevens's equation.

The equality of the slopes of the regression between the different assessors can also be tested.

9.5 Rescaling methods

9.5.1 Total rescaling

The reasoning on which this method is based is as follows. Since each assessor has evaluated the same set of samples, the total magnitude of the response for this set of samples should be identical for each assessor. Therefore, the scale for each assessor is brought to the same total magnitude for a set of common samples.

The procedure is as follows.

For all of the samples evaluated by all of the assessors:

— calculate the mean of the logarithm of the estimations of each assessor;

— calculate the general mean for all the assessors.

For each assessor:

— calculate the correction value which, once added to its mean, will make it equal to the mean of the group;

— add its correction value to all the estimations of each assessor.

An example of carrying out total rescaling is given in B.2.

9.5.2 Rescaling in relation to the reference sample

If one or more reference samples, randomly coded, have been incorporated into the test series, first calculate for each assessor the mean of the estimations relating to the reference samples [first sample and hidden reference(s), if any]. Then, calculate the correction value that would bring this mean to a fixed value. In order to rescale the data obtained for the test samples, multiply each estimation of an assessor by the correction value calculated from the reference sample(s).

It is advisable to note that the global ANOVA, as well as the procedure for total rescaling, gives rise to a smaller mean square error than the procedure of rescaling in relation to the reference sample. As indicated in 8.2, the reference sample should have an intensity close to the geometric mean of all the samples for the whole panel. It has been shown that the error is lower when the intensity of the reference sample is equal to the geometric mean[6]. The closer the value for the reference sample is to the true geometric mean, the better.

9.5.3 External rescaling

Various forms of external rescaling have been reported in the literature. After evaluating the test samples, the assessor receives a verbal response scale comprising between four and eleven graduations. It consists of expressions such as:

— extremely intense;

— very intense;

— moderately intense;

— slightly intense, etc.

The panel leader requests the assessor to assign magnitude estimations to these expressions in a manner that is consistent with the scale being used when evaluating the test samples. Results given by each assessor are rescaled using a correction value calculated by applying the total rescaling method to the values assigned to the expressions of the verbal response scale.

An example of external rescaling is given in B.4.

10 Test report

The test report shall indicate:

— the objectives of the study;

— the test results;

— the number of samples and a description of the samples;

— recourse, if any, to a reference sample and, if used, the nature of this sample;

— replication, if any, of the tests;

— the number of assessors and their level of qualification;

— the general conditions of testing, such as test environment, date and time;

— any other information allowing the overall validity of the tests to be evaluated;

— the reference to the number of this document, i.e. ISO 11056, together with an indication of adjustments, if any, made to the method;

— the name of the person in charge of the assay.

— the date of the test.

Annex A
(informative)

Questionnaire models

A.1 Questionnaire model without fixed modulus for the reference sample

Workplace code: ..... Date: .....

1. A reference sample of orange juice coded "R" is presented to you.

You are requested to taste it and to rate the intensity of its acid taste with the aid of a number of your choice:

Response:

Memorize well the intensity of its acidity.

2. Six orange drinks are presented to you.

You are requested to evaluate them in the order given below.

For each of the samples, assign a value to the intensity of the acid taste, in proportion to the value of the reference "R".

You have to retaste the reference before each sample.

Sample 561

Sample 274

Sample 935

Sample 803

Sample 417

Sample 127

A.2 Questionnaire model with fixed modulus for the reference sample

Workplace code: ..... Date: .....

1. A reference sample of orange juice coded "R" is presented to you.

The value assigned to its "acidity" is equal to 50.

Taste this sample and memorize the intensity of its acidity.

2. Six orange drinks are presented to you.

You are requested to evaluate them in the order indicated below.

For each of the samples, assign a value to the intensity of the acid taste, in proportion to the value (50) given to the reference "R".

You have to retaste the reference before each sample.

Sample 561

Sample 274

Sample 935

Sample 803

Sample 417

Sample 127

Annex B
(informative)

Examples of data analysis

B.1 Data analysis of an assay without replication and without rescaled data — All assessors once evaluated all products in the series

B.1.1 Case presented

Table B.1 presents the results obtained by seven experienced assessors having evaluated the intensity of the bitterness of six samples of a beverage containing various quantities of caffeine. For Assessors 1, 2 and 3, sample 274 is the external reference with an assigned numerical value equal to 20. For Assessors 5, 6 and 7, sample 803 is the external reference with an assigned value equal to 40. Finally, for Assessor 4, sample 935 is the reference with a value equal to 32. For each assessor, the external reference was the first sample in the series; the other five samples were presented in a random order that was different depending on the assessor. The assessors did not evaluate the external reference (the scores in Table B.1 are those assigned by the panel leader), but they tasted this reference at least five times (once before tasting each of the five other products) by associating it with the score given by the panel leader.

Natural logarithms have been calculated with three digits and are presented in Table B.1 in parentheses.

Table B.1 — Table of data relative to the six samples

Assessor	Treatment codes					
	561	274	935	803	417	127
	Concentrations (mg/100 ml)					
	9	18	36	40	72	144
	Magnitude estimations (logarithms)					
1	10 (2,303)	20 (2,996)	35 (3,555)	40 (3,689)	70 (4,248)	140 (4,942)
2	8 (2,079)	20 (2,996)	38 (3,638)	44 (3,784)	85 (4,443)	160 (5,075)
3	8 (2,079)	20 (2,996)	36 (3,584)	40 (3,689)	75 (4,317)	150 (5,010)
4	7 (1,946)	15 (2,708)	32 (3,466)	37 (3,611)	70 (4,248)	135 (4,905)
5	12 (2,485)	25 (3,219)	38 (3,638)	40 (3,689)	75 (4,317)	145 (4,977)
6	12 (2,485)	22 (3,091)	35 (3,555)	40 (3,689)	80 (4,382)	160 (5,075)
7	9 (2,197)	18 (2,890)	35 (3,555)	40 (3,689)	74 (4,304)	145 (4,977)
Mean log	2,225	2,985	3,570	3,691	4,323	4,994

This file is in ME(2019) folder (USB DISK F) under the name: Table Annex B1. It is imported into R with the following three commands:

```
tableb1<- read.table ("F:/ME(2019)/Table Annex B1.txt", header = T, sep="\t", dec=".")
attach(tableb1)
names(tableb1)
```

The file tableb1 has 7 columns (Assessor, Treatment, Score, LogScore, LogScoreresc, Conc LogConc) and 42 rows (6 treatments × 7 assessors). The command:

```
round(with(tableb1,tapply(LogScore, list(Treatment),mean)), digits=3)
```

leads to the following values of the row "mean log" in Table B.1:

1T	2T	3T	4T	5T	6T
2,225	2,985	3,570	3,691	4,323	4,995

B.1.2 Determination of the existence of significant differences

A two-way ANOVA was applied to the logarithms of Table B.1 using the R command:

```
summary(aov(LogScore ~ Assessor +Treatment, data= tableb1))
```

The results are given in Table B.2. This table shows that the effect induced by the Treatment factor is very significant, which is logical, given the very large differences in caffeine concentration (6, 18, 36, etc.).

Table B.2 — Results of two-way ANOVA of Table B.1

Source of variation	Degrees of freedom	Sum of squares	Mean square	F value	$P > F$
Assessor	6	0,24	0,040	4,55	0,002
Treatment	5	33,18	6,635	753,11	< 2e-16
Residuals	30	0,26	0,009	—	—

Multiple comparison test: Tukey's test is one of the numerous multiple comparison tests likely to be used for determining which samples differ significantly from one another. In this test, the least significant difference (LSD) is determined as shown by Formula (B.1):

$$C\times\sqrt{\frac{1}{2}s^2\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \tag{B.1}$$

where

s^2 is the mean square of the row "residuals" in Table B.2;

n_i is the number of observations used for the calculation of the first mean;

n_j is the number of observations used for the calculation of the second mean;

C is a factor; a function of the degrees of freedom of the row "residuals", of the total number of treatments, and of the α-risk chosen; its value is given, for example, in Table B.6. Critical values of Tukey's studentized range in Reference [1].

In this example (6 treatments and 30 degrees of freedom for residuals), C is equal to 4,30 for an α-risk = 0,05 and the LSD is equal to Formula (B.2):

$$4{,}30\times\sqrt{\left(0{,}009\times0{,}5\times\left(\frac{1}{7}+\frac{1}{7}\right)\right)}=0{,}154 \tag{B.2}$$

The only two samples not differing in any significant manner (see the last row in Table B.1) are 803 and 935. Their means differ by only 0,121. This conclusion is logical. The caffeine concentrations of samples 803 and 935 are close: they only differ (in log) of 0,046, whereas the difference of caffeine concentration (in log): between the other pairs of adjacent treatments is equal to 0,255 (treatment 417 versus 803) and to 0,301 (treatment 274 versus 561, treatment 935 versus 274, treatment 127 versus 417).

The calculation above can be obtained by using the TukeyHSD command of R:

```
TukeyHSD(aov(LogScore ~ Assessor + Treatment, data= b1), "Treatment")
```

This command leads to:

```
$Treatment
```

	diff	lwr	upr	p adj
2T-1T	0.7602857	0.60768686	0.9128846	0.0000000
3T-1T	1.3452857	1.19268686	1.4978846	0.0000000
4T-1T	1.4665714	1.31397258	1.6191703	0.0000000
5T-1T	2.0978571	1.94525829	2.2504560	0.0000000
6T-1T	2.7695714	2.61697258	2.9221703	0.0000000
3T-2T	0.5850000	0.43240115	0.7375988	0.0000000
4T-2T	0.7062857	0.55368686	0.8588846	0.0000000
5T-2T	1.3375714	1.18497258	1.4901703	0.0000000
6T-2T	2.0092857	1.85668686	2.1618846	0.0000000
4T-3T	0.1212857	−0.03131314	0.2738846	0.1824317
5T-3T	0.7525714	0.59997258	0.9051703	0.0000000
6T-3T	1.4242857	1.27168686	1.5768846	0.0000000
5T-4T	0.6312857	0.47868686	0.7838846	0.0000000
6T-4T	1.3030000	1.15040115	1.4555988	0.0000000
6T-5T	0.6717143	0.51911544	0.8243131	0.0000000

For each pair of treatments, the command gives:

a) the difference between the two treatments;

b) the lower limit (lwr) and the upper limit (upr) of confidence for 95 %;

c) the *p* value (the word “adj” *is* useless).

The only two samples not differing in any significant manner (at α = 0,05) are the 4T-3T treatments (803 and 935).

NOTE 1 When the Assessor factor is considered as a random factor, it is necessary to use the lmer() command (from lmerTest package R). The commands are:

```
library(lmerTest)
resb1 <-lmer(LogScore ~Treatment +(1|Assessor), data = tableb1)
anova(resb1)
ranova(resb1)
```

The results are identical for the Treatment factor (the *p* value is identical to that of Table B.2) because the experimental design is balanced. Furthermore, the *p* value is different for the Assessor factor: it is equal to 0,006. This value is higher than that of Table B.2: 0,002. This observation was expected.

The Assessor effect is usually lower (and therefore the *p* value is higher) when the Assessor factor is considered random rather than fixed.

NOTE 2 It is possible to treat the data of Table B.1 with rescaling performed on the six scores of each assessor as indicated in B.2. In tableb1, the values obtained after rescaling are given in the column "LogScoreresc R command":

```
summary(aov(LogScoreresc ~ Assessor +Treatment, data = tableb1))
```

This leads to Table B.3. The rescaling leads to an identical sum for all seven assessors. It is equal to 3,631 and the sum of squares is equal to 0 for the row "assessor".

Table B.3 — Results of two-way ANOVA of Table B.1 after rescaling

Source of variation	Degrees of freedom	Sum of squares	Mean square	*F* value	*Pr* (> *F*)
Assessor	6	0,00	0,000	0	1
Treatment	5	33,17	6,635	739	< 2 e-16
Residuals	30	0,26	0,009	—	—

B.2 Data analysis of an assay without replication but with internal rescaling — All assessors once evaluated all products in the series

Table B.4 indicates the results of a study in which panellists gave values of perceived sweetness to a series of sucrose solutions. A solution of 2 % of sucrose was used as a reference sample which was assigned a score of 10 by all the assessors. However, the scores of the reference sample were not included in the data analysis. Reference sample 2 % was also rated as a hidden sample.

Table B.4 — Table of data relative to the six samples of sucrose

% sucrose concentration (mass/volume)	Assessors					
	L.M.	B.W.	M.M.	H.H.	J.D.	J.F.
0,5	5	1	1	1	2	0,5
1	1	4	2,5	1	2	2
2	10	10	8	5	10	5
4	20	15	20	30	20	25
8	50	30	40	100	25	50
16	100	50	80	300	40	200
Geometric mean	13,08	9,83	10,42	12,85	9,63	10,38

NOTE The geometric mean of each assessor is given by:

$$GM = \sqrt[n]{(x_1)\times(x_2)\ldots(x_n)}$$

For example, for panellist L.M.:

$$GM = \sqrt[6]{(5\times1\times10\times20\times50\times100)} = 13{,}08$$

Table B.4 leads to Table B.5 after logarithmic transformation.

Table B.5 — Table B.4 after logarithmic transformation

	L.M.	B.W.	M.M.	H.H.	J.D.	J.F.	
0,5 (−0,693)	1,609	0	0	0	0,693	−0,693	
1 (0,000)	0	1,386	0,916	0	0,693	0,693	
2 (0,693)	2,303	2,303	2,079	1,609	2,303	1,609	
4 (1,386)	2,996	2,708	2,996	3,401	2,996	3,219	
8 (2,079)	3,912	3,401	3,689	4,605	3,219	3,912	
16 (2,773)	4,605	3,912	4,382	5,704	3,689	5,298	
Mean (log)	2,571	2,285	2,344	2,553	2,266	2,34	2,393
Correction factor	−0,178	0,108	0,049	−0,16	0,127	0,053	

The general mean of Table B.5 is equal to 2,393. For each assessor, the correction factor allows to rescale the obtained values. This factor is given by: general mean – assessor's mean. For example, for assessor L.M., it is equal to 2,393 – 2,571 = −0,178. Hence, 1,609 – 0,178 = 1,431 for the concentration 0,5; 0 – 0,178 = −0,178 for concentration 1, 2,303 – 0,178 = 2,125 for concentration 2; etc.

The calculations lead to Table B.6.

Table B.6 — Table B.4 after total rescaling on log values

% sucrose concentration (mass/volume, in log)	Assessors						
	L.M.	B.W.	M.M.	H.H.	J.D.	J.F.	Mean
−0,693	1,431	0,108	0,049	−0,160	0,820	−0,640	0,268
0,000	−0,178	1,494	0,965	−0,160	0,820	0,746	0,614
0,693	2,125	2,411	2,128	1,449	2,43	1,662	2,034
1,386	2,818	2,816	3,045	3,241	3,123	3,272	3,053
2,079	3,734	3,509	3,738	4,445	3,346	3,965	3,790
2,773	4,427	4,020	4,431	5,544	3,816	5,351	4,598
Mean (log)	2,393	2,393	2,393	2,393	2,393	2,393	—

This file is in ME(2019) folder (USB DISK F) under the name: Table Annex B2. It is imported into R with the following three commands:

```
tableb6 < - read.table ("F:/ME(2019)/Table Annex B2.txt", header = T, sep="\t", dec=".")
attach(tableb6)
names(tableb6)
```

The file tableb6 has 7 columns (Assessor, Concsu, Conc, LogConc, Score, LogScore, LogScoreresc) and 36 rows (6 assessors × 6 conc). The command:

```
round(with(tableb6,tapply(LogScoreresc, list(Conc),mean)), digits=3)
```

leads to the following values of the "mean" column in Table B.6:

1c	2c	3c	4c	5c	6c
0,268	0,614	2,034	3,053	3,789	4,598

The assessors have the same mean(log): 2,393. It is given with command:

```
round(with(tableb6,tapply(LogScoreresc, list(Assessor),mean)), digits=3)
```

A two-way ANOVA was applied to the data of Table B.6 using the R command:

```
summary(aov(LogScoreresc ~ Assessor+Conc, data= tableb6))
```

The results are given in Table B.7. The sum of square of the Assessor factor is equal to 0 since the data has been completely rescaled.

Table B.7 — Results of two-way ANOVA of Table B.6

Source of variation	Degrees of freedom	Sum of squares	Mean square	*F* value	*Pr* > *F*
Assessor	5	0	0	—	—
Conc	5	90,33	18,066	49,5	3,62 e-12
Residuals	25	9,12	0,365	—	—

The Conc factor is very highly significant. R command:

```
TukeyHSD(aov(LogScoreresc ~ Assessor + Conc, data= tableb6), "Conc")
```

shows that the products of pairs with adjacent concentrations 2c-1c and 3c-2c are perceived as significantly different (their *p* values are < 0,05). But the products of pairs 4c-3c, 5c-4c, 6c-5c are not perceived as significantly different (their *p* values are equal to 0,07, 0,31 and 0,22, respectively).

This command leads to:

```
              diff         lwr       upr         p adj
2c-1c    0.3465000 -0.72844156 1.1421442 0.9157442
3c-1c    1.7661667  0.69122510 2.841108  0.0004102
4c-1c    2.7845000  1.70955844 3.859442  0.0000003
5c-1c    3.5215000  2.44655844 4.596442  0.0000000
6c-1c    4.3301667  3.25522510 5.405108  0.0000000
3c-2c    1.4196667  0.34472510 2.494608  0.0049459
4c-2c    2.4380000  1.36305844 3.512942  0.0000035
5c-2c    3.1750000  2.10005844 4.249942  0.0000000
6c-2c    3.9836667  2.90872510 5.058608  0.0000000
4c-3c    1.0183333  0.05660823 2.093275  0.0707337
5c-3c    1.7553333  0.68039177 2.830275  0.0004437
6c-3c    2.5640000  1.48905844 3.638942  0.0000015
5c-4c    0.7370000 -0.33794156 1.811942  0.3129969
6c-4c    1.5456667  0.47072510 2.620608  0.0020171
6c-5c    0.8086667 -0.26627490 1.883608  0.2240927
```

B.3 Data analysis in cases of internal rescaling on a subgroup of samples

Suppose that in the example of Table B.1 all assessors did not evaluate all samples in the series. Assessors 2, 3, 4, 5, 6 and 7 evaluated only six samples. The experimental design is not balanced.

In fact, Table B.4 is identical to Table B.1 except that the scores of Assessors 2, 4 and 6 given to treatment 274 and the scores of Assessors 3, 5 and 7 given to treatment 417 have been deleted.

As all the assessors evaluated samples 561, 935, 803 and 127, the rescaling can be carried out on this sub-group of samples as follows:

a) determine the rescaling factors by calculating initially for each assessor the mean logarithm for the four common samples (e.g. for Assessor 1: (2,303 + 3,555 + 3,689 + 4,942)/4 = 3,622, see Table B.8);

b) next, calculate the mean logarithm of the estimations for the whole panel (it is equal to 3,620);

c) next, calculate the correction value for each assessor by subtracting the mean per assessor from the mean of the group (e.g. for Assessor 1: 3,620 – 3,622 = –0,002);

d) finally, use this value to correct each assessor's scores and obtain Table B.9.

Table B.8 — Logarithms of the estimations provided by the assessors and calculation of the correction factor

Assessor	Treatment codes						Mean of the common sub-group	Correction factor
	561	274	935	803	417	127		
	Magnitude estimations (logarithms)							
1	2,303	2,996	3,555	3,689	4,248	4,942	3,622	–0,002
2	2,079	—	3,638	3,784	4,443	5,075	3,644	–0,024
3	2,079	2,996	3,584	3,689	—	5,011	3,591	+0,029
4	1,946	—	3,466	3,611	4,248	4,905	3,482	+0,138
5	2,485	3,219	3,638	3,689	—	4,977	3,697	–0,077
6	2,485	—	3,555	3,689	4,382	5,075	3,701	–0,081
7	2,197	2,890	3,555	3,689	—	4,977	3,604	+0,016
Group	—	—	—	—	—	—	3,620	—

Table B.9 — Log (estimations) after rescaling

Assessor	Treatment codes					
	561	274	935	803	417	127
	Magnitude estimations (logarithms)					
1	2,301	2,994	3,553	3,687	4,246	4,940
2	2,055	—	3,614	3,760	4,419	5,051
3	2,108	3,025	3,613	3,718	—	5,040
4	2,084	—	3,604	3,749	4,386	5,043
5	2,408	3,142	3,561	3,612	—	4,900
6	2,404	—	3,474	3,608	4,301	4,994
7	2,213	2,906	3,571	3,705	—	4,993
Mean log	2,225	3,017	3,570	3,691	4,338	4,994

This file is in ME(2019) folder (USB DISK F) under the name: Table Annex B3. It is imported into R with the following three commands:

```
tableb9<- read.table ("F:/ME(2019)/Table Annex B3.txt", header = T, sep="\t", dec=".")
attach(tableb9)
names(tableb9)
```

The file tableb9 has 5 columns (Assessor, Treatment, Treatmentbis, LogScore, LogScoreresc) and 36 rows (6 rows brought by Assessor 1 and 30 rows brought by Assessors 2, 3, 4, 5, 6 and 7; each of these assessors bringing 5 rows). The R command:

```
round(with(tableb9, tapply(LogScoreresc, list (Treatment), mean)), digits=3)
```

leads to the following values of the row "mean log" in Table B.9:

1T	2T	3T	4T	5T	6T
2,225	3,017	3,570	3,691	4,338	4,994

An ANOVA was applied to the data of Table B.9 using the R command:

```
summary(aov(LogScoreresc ~ Treatment + Assessor, data = tableb9))
```

The results are given in Table B.10.

Table B.10 — Results of ANOVA of Table B.9

Source of variation	Degrees of freedom	Sum of squares	Mean square	*F* value	*Pr* > *F*
Treatment	5	30,416	6,083	641,543	< 2 e-16
Assessor	6	0,010	0,002	0,173	0,982
Residuals	24	0,228	0,009	—	—

The Treatment factor is very highly significant. As the rescaling was performed only on four treatments, the sum of squares of the Assessor factor is not equal to 0. The means of each assessor given by the R command:

```
round(with(tableb9, tapply(LogScoreresc, list (Assessor), mean)), digits=3)
```

are:

As1	As2	As3	As4	As5	As6
3.620	3.780	3.501	3.773	3.756	3.478

Multiple comparison test: For the calculation of the LSD, it is necessary to take into account the fact that the means of the different treatments are not obtained with the same number of observations (four for 274, 417 and 121, seven for 567, 935, 803). The Tukey-Kramer method (see Reference [2]) may be used. In this example (6 treatments and 24 degrees of freedom for residuals), C is equal to 4,37 for an α-risk = 0,05.

Thus, the LSD for the pair (274, 417), which concerns treatments for which four measurements are available, is equal to:

$$4{,}37\sqrt{0{,}009\times 0{,}5\left(\frac{1}{4}+\frac{1}{4}\right)}=0{,}207$$

For the pair (561, 274), LSD is equal to:

$$4{,}37\sqrt{0{,}009\times 0{,}5\left(\frac{1}{7}+\frac{1}{4}\right)}=0{,}184$$

Finally, for the pair (561, 935), LSD is equal to:

$$4{,}37\sqrt{0{,}009\times 0{,}5\left(\frac{1}{7}+\frac{1}{7}\right)}=0{,}157$$

Thus, only treatments 803 and 935 are not significantly different. This conclusion is identical to that of B.1.

The previous calculations can be directly obtained by using the R command:

```
TukeyHSD(aov(LogScoreresc~ Treatment+Assessor, data = tableb9), "Treatment")
```

This command leads to:

	diff	lwr	upr	p adj
2T-1T	0.7920357	0.60332414	0.9807473	0.0000000
3T-1T	1.3452857	1.18435195	1.5062195	0.0000000
4T-1T	1.4665714	1.30563766	1.6275052	0.0000000
5T-1T	2.1132857	1.92457414	2.3019973	0.0000000
6T-1T	2.7697143	2.60878052	2.9306481	0.0000000
3T-2T	0.5532500	0.36453843	0.7419616	0.0000000
4T-2T	0.6745357	0.48582414	0.8632473	0.0000000
5T-2T	1.3212500	1.10835463	1.5341454	0.0000000
6T-2T	1.9776786	1.78896700	2.1663901	0.0000000
4T-3T	0.1212857	−0.03964805	0.2822195	0.2210445
5T-3T	0.7680000	0.57928843	0.9567116	0.0000000
6T-3T	1.4244286	1.26349480	1.5853623	0.0000000
5T-4T	0.6467143	0.45800271	0.8354259	0.0000000
6T-4T	1.3031429	1.14220909	1.4640766	0.0000000
6T-5T	0.6564286	0.46771700	0.8451401	0.0000000

NOTE When the Assessor factor is considered to be random, the function lmer() (from Package lmerTest of R):

```
library(lmerTest)
resb9 -lmer(LogScoreresc ~Treatment +(1|Assessor), data = tableb9)
anova(resb9)
ranova(resb9)
```

leads for the Treatment factor to an *F* value equal to 768,74 and to a *p* value equal to 2,2 e-16. For the Assessor factor, the *p* value is equal to 1: the differences between assessors are zero.

B.4 Data analysis in cases of external rescaling

After carrying out the main experiment, the assessors assigned magnitude values to a verbal calibration scale. To illustrate this, a five-point scale ranging from “extremely bitter” to “slightly bitter” has been treated. The panel leader asked the assessor to assign magnitude estimations to these expressions in a manner consistent with the scale being used when evaluating the test samples. The hypothetical results for this exercise are presented in Table B.11.

Table B.11 — Hypothetical results of ratings for the verbal scale

Assessor	Slightly bitter	Bitter	Moderately bitter	Very bitter	Extremely bitter	Rescaling factor[a]
1	5	25	50	100	150	−0,002 5
2	5	30	60	100	160	−0,088 0
3	5	25	50	100	150	−0,002 5
4	5	20	45	90	140	+0,098 0
5	5	25	50	100	150	−0,002 5
6	3	30	55	110	170	0
7	5	25	50	100	150	−0,002 5

[a] Calculated as in B.2.

The data in Table B.12 were used as follows to obtain the rescaled values of Table B.13:

— first, calculate the correction values using the total rescaling method applied to the calibration scores;

— next, correct the data of Table B.8 by adding to the data of each assessor his/her correction value.

Note that, in Table B.12, only Assessor 1 evaluated all six products in the series. The other assessors evaluated only five samples (e.g. Assessor 2 did not evaluate treatment 274). The experimental design is therefore not balanced. Note also that the empty cells in Table B.11 are not the same than those of Table B.8. For example, for treatment 561, there is an empty cell in Table B.8 (that of Assessor 2) but there are no empty cells in Table B.11.

Table B.12 — Log (estimations) given by the assessors

Assessor	Treatment codes					
	561	274	935	803	417	127
1	2,303	2,996	3,555	3,689	4,248	4,942
2	—	2,996	3,638	3,784	4,443	5,075
3	2,079	—	3,584	3,689	4,317	5,011
4	1,946	2,708	—	3,611	4,248	4,905
5	2,485	3,219	3,638	—	4,317	4,977
6	2,485	3,091	3,555	3,689	—	5,075
7	2,197	2,890	3,555	3,689	4,304	—

Table B.13 — Log (estimations) of Table B.12 after rescaling

Assessor	Treatment codes					
	561	274	935	803	417	127
1	2,300	2,993	3,553	3,686	4,246	4,939
2	—	2,908	3,550	3,696	4,355	4,987
3	2,077	—	3,581	3,686	4,315	5,008
4	2,044	2,806	—	3,709	4,346	5,003
5	2,482	3,216	3,635	—	4,315	4,974
6	2,485	3,091	3,555	3,689	—	5,075
7	2,195	2,888	3,553	3,686	4,302	—
Group	2,264	2,984	3,571	3,692	4,313	4,998

This file is in ME(2019) folder (USB DISK F) under the name Table Annex B4. It is imported into R with the following three commands:

```
tableb13<- read.table("F:/ME(2019)/Table Annex B4.txt", header = T, sep="\t", dec=".")
attach(tableb13)
names(tableb13)
```

The file tableb13 has 4 columns (Assessor, Treatment, Score, LogScoreresc) and 36 rows (6 rows brought by Assessor 1 and 30 rows brought by Assessors 2, 3, 4, 5, 6 and 7). The command:

```
round(with(tableb13,tapply(LogScoreresc, list(Treatment),mean)), digits=3)
```

leads to the following values of the row "group" in Table B.13:

1T	2T	3T	4T	5T	6T
2,264	2,984	3,571	3,692	4,313	4998

An ANOVA was performed with the R command:

```
summary(aov(LogScoreresc ~ Treatment + Assessor, data = tableb13))
```

The results are given in Table B.14. As the design is unbalanced, the sum of squares of the Assessor factor is not equal to 0. The Treatment factor is very highly significant.

Table B.14 — Results of ANOVA of Table B.13

Source of variation	Degrees of freedom	Sum of squares	Mean square	*F* value	*Pr* > *F*
Treatment	5	27,770	5,554	669,077	2 < e-16
Assessor	6	0,122	0,020	2,455	0,054
Residuals	24	0,199	0,008	—	—

The LSD calculated as described in B.1.2 for 6 treatments with 6 observations by treatment, 24 degrees of freedom for the row "residuals" and an α-risk = 0,05 is equal to:

$$4{,}37\times\sqrt{0{,}008\times0{,}5\left(\frac{1}{6}+\frac{1}{6}\right)}=0{,}160$$

The only treatments which do not differ significantly are 935 and 803. This conclusion is identical to that of B.1 and B.3.

That results are also given by R command:

```
TukeyHSD(aov(LogScoreresc~ Treatment+Assessor, tableb13), "Treatment")
```

This command leads to:

```
Tukey multiple comparisons of means
    95 % family-wise confidence level
Fit: aov(formula = LogScoreresc ~ Treatment + Assessor, data = tableb13)


$Treatment  i
                    diff           lwr           upr          p adj
    2T-1T       0.7198333       0.5571896       0.8824771       0.000000
    3T-1T       1.3073333       1.1446896       1.4699771       0.000000
    4T-1T       1.4281667       1.2655229       1.5908104       0.000000
    5T-1T       2.0493333       1.8866896       2.2119771       0.000000
    6T-1T       2.7338333       2.5711896       2.8964771       0.000000
    3T-2T       0.5875000       0.4248563       0.7501437       0.000000
    4T-2T       0.7083333       0.5456896       0.8709771       0.000000
    5T-2T       1.3295000       1.1668563       1.4921437       0.000000
    6T-2T       2.0140000       1.8513563       2.1766437       0.000000
    4T-3T       0.1208333      -0.0418104       0.2834771       0.233788
    5T-3T       0.7420000       0.5793563       0.9046437       0.000000
    6T-3T       1.4265000       1.2638563       1.5891437       0.000000
    5T-4T       0.6211667       0.4585229       0.7838104       0.000000
    6T-4T       1.3056667       1.1430229       1.4683104       0.000000
    6T-5T       0.6845000       0.5218563       0.8471437       0.000000
```

The two adjacent treatments that do not differ significantly are 4T (935) and 3T (803).

NOTE 1 As the design is unbalanced, the order Assessor + Treatment in the function aov(LogScoreresc...) gives different results from those of Treatment Assessor. With the order Assessor + Treatment, the *p* value of the Assessor factor is < 0,05. But the order Treatment – Assessor is usually considered be the most relevant order.

NOTE 2 When the Assessor factor is considered as a random factor, R commands:

```
library(lmerTest)
resb13 <-(lmer(LogScoreresc~ Treatment +(1|Assessor), data = tableb13))
anova(resb13)
ranova(resb13)
```

lead for the Treatment factor to an *F* value equal to 655,32 and to a *p* value equal to 2,2 e-16. For the Assessor factor, the *p* value is equal to 0,14: the differences between assessors are not significant at α = 0,05 (the *p* value was equal to 0,054 when the Assessor factor was considered fixed).

B.5 Data analysis in cases where there are replications

B.5.1 General

Table B.15 presents a list of the replicated magnitude estimations assigned by the same members of the panel as those of which the results are given in Table B.1. After completing the evaluation of the first replication samples, the assessors immediately evaluated the second replication samples, but in a different order from the first replication. The assessors were not informed that the evaluation included two repetitions.

Table B.15 — Scores for replications 1 and 2

Assessor	Replicate	Treatment codes					
		561	274	935	803	417	127
		Magnitude estimations (logarithms)					
1	1	10 (2,303)	20 (2,996)	35 (3,555)	40 (3,689)	70 (4,248)	140(4,942)
	2	15 (2,708)	25 (3,219)	35 (3,555)	38 (3,638)	70 (4,248)	135 (4,905)
2	1	8 (2,079)	20 (2,996)	38 (3,638)	44 (3,784)	85 (4,443)	160 (5,075)
	2	8 (2,079)	15 (2,708)	35 (3,555)	45 (3,807)	90 (4,500)	180 (5,193)
3	1	8 (2,079)	20 (2,996)	36 (3,584)	40 (3,689)	75 (4,317)	150(5,011)
	2	10 (2,303)	20 (2,996)	35 (3,555)	35 (3,555)	70 (4,248)	145 (4,977)
4	1	7 (1,946)	15 (2,708)	32 (3,466)	37 (3,611)	70 (4,248)	135 (4,905)
	2	10 (2,303)	20 (2,996)	35 (3,555)	38 (3,638)	65 (4,174)	130 (4,868)
5	1	12 (2,485)	25 (3,219)	38 (3,638)	40 (3,689)	75 (4,317)	145 (4,977)
	2	10 (2,303)	25 (3,219)	35 (3,555)	40 (3,689)	80 (4,382)	150 (5,011)
6	1	12 (2,485)	22 (3,091)	35 (3,555)	40 (3,689)	80 (4,382)	160 (5,075)
	2	10 (2,303)	20 (2,996)	35 (3,555)	40 (3,689)	80 (4,382)	160 (5,075)
7	1	9 (2,197)	18 (2,890)	35 (3,555)	40 (3,689)	74 (4,304)	145 (4,977)
	2	10 (2,303)	15 (2,708)	35 (3,555)	38 (3,638)	70 (4,248)	140 (4,942)
Mean log	—	2,277	2,981	3,563	3,678	4,317	4,995

This file is in ME(2019) folder (USB DISK F) under the name Table Annex B15. It is imported into R with the following three commands:

```
tableb15<- read.table("F:/ME(2019)/Table Annex B5.txt", header = T, sep="\t", dec=".")
attach(tableb15)
names(tableb15)
```

The file tableb15 has 7 columns (Assessor, Treatment, Score, LogScore, Replicate, Conc, LogConc) and 84 rows (7 assessors × 6 treatments × 2 replicates). The 12 scores of Assessor 1 are in rows 1 to 12, the 12 scores of Assessor 2 are in rows 13 to 24, etc. The command:

```
round(with(tableb15,tapply(LogScore, list(Treatment),mean)), digits=3)
```

leads to the following values of the row “mean log” in Table B.15.

1T	2T	3T	4T	5T	6T
2,277	2,981	3,563	3,678	4,317	4,995