INTERNATIONAL STANDARD

**ISO 13528**

Third edition
2022-08

# Statistical methods for use in proficiency testing by interlaboratory comparison

*Méthodes statistiques utilisées dans les essais d'aptitude par comparaison interlaboratoires*

Reference number
ISO 13528:2022(E)

© ISO 2022



## COPYRIGHT PROTECTED DOCUMENT

© ISO 2022

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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



© ISO 2022 – All rights reserved

# Contents

Page

**Foreword** ..... **v**

**0 Introduction** ..... **vi**

**1 Scope** ..... **1**

**2 Normative references** ..... **1**

**3 Terms and definitions** ..... **1**

**4 General principles** ..... **4**
4.1 General requirements for statistical methods ..... 4
4.2 Basic model ..... 5
4.3 General approaches for the evaluation of performance ..... 5

**5 Guidelines for the statistical design of proficiency testing schemes** ..... **6**
5.1 Introduction to the statistical design of proficiency testing schemes ..... 6
5.2 Basis of a statistical design ..... 6
5.3 Considerations for the statistical distribution of results ..... 7
5.4 Considerations for small numbers of participants ..... 8
5.5 Guidelines for choosing the reporting format ..... 8
5.5.1 General requirements for reporting format ..... 8
5.5.2 Reporting of replicate measurements ..... 9
5.5.3 Reporting of 'less than' or 'greater than' a limit (censored data) ..... 9
5.5.4 Number of significant digits ..... 9

**6 Guidelines for the initial review of proficiency testing items and results** ..... **10**
6.1 Homogeneity and stability of proficiency test items ..... 10
6.2 Considerations for different measurement methods ..... 11
6.3 Blunder removal ..... 11
6.4 Visual review of data ..... 12
6.5 Robust statistical methods ..... 12
6.6 Outlier techniques for individual results ..... 13

**7 Determination of the assigned value and its standard uncertainty** ..... **14**
7.1 Choice of method of determining the assigned value ..... 14
7.2 Determining the uncertainty of the assigned value ..... 14
7.3 Formulation ..... 15
7.4 Certified reference material ..... 16
7.5 Results from one laboratory ..... 16
7.6 Consensus value from expert laboratories ..... 17
7.7 Consensus value from participant results ..... 18
7.8 Comparison of the assigned value with an independent reference value ..... 19

**8 Determination of criteria for evaluation of performance** ..... **20**
8.1 Approaches for determining evaluation criteria ..... 20
8.2 By perception of experts ..... 21
8.3 By experience from previous rounds of a proficiency testing scheme ..... 21
8.4 By use of a general model ..... 21
8.5 Using the repeatability and reproducibility standard deviations from a previous collaborative study of precision of a measurement method ..... 22
8.6 From data obtained in the same round of a proficiency testing scheme ..... 22
8.7 Monitoring interlaboratory agreement ..... 23

**9 Calculation of performance statistics** ..... **24**
9.1 General considerations for determining performance ..... 24
9.2 Limiting the uncertainty of the assigned value ..... 24
9.3 Estimates of deviation (measurement error) ..... 25
9.4 $z$ scores ..... 26
9.5 $z'$ scores ..... 27

© ISO 2022 – All rights reserved

9.6 Zeta scores ($\zeta$) 28
9.7 $E_n$ scores 29
9.8 Evaluation of participant uncertainties in testing 30
9.9 Combined performance scores 31

**10 Graphical methods for describing performance scores 32**
10.1 Application of graphical methods 32
10.2 Histograms of results or performance scores 32
10.3 Kernel density plots 33
10.4 Bar-plots of standardized performance scores 34
10.5 Youden plot 34
10.6 Plots of repeatability standard deviations 35
10.7 Split samples 36
10.8 Graphical methods for combining performance scores over several rounds of a proficiency testing scheme 37

**11 Design and analysis of qualitative proficiency testing schemes (including nominal and ordinal properties) 38**
11.1 Types of qualitative data 38
11.2 Statistical design 38
11.3 Assigned values for qualitative proficiency testing schemes 39
11.4 Performance evaluation and scoring for qualitative proficiency testing schemes 40

**Annex A** (normative) **Symbols 42**

**Annex B** (informative) **Homogeneity and stability of proficiency test items 44**

**Annex C** (informative) **Robust analysis 52**

**Annex D** (informative) **Additional guidance on statistical procedures 63**

**Annex E** (informative) **Illustrative examples 68**

**Annex F** (Informative) **Example of computer code for plotting and resampling analysis ("bootstrapping") of PT results 91**

**Bibliography 92**


© ISO 2022 – All rights reserved

# 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 on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

This third edition of ISO 13528 cancels and replaces the second edition (ISO 13528:2015), of which it constitutes an minor revision. The changes are as follows:

— notes have been added to 10.1, 10.4.3 and 10.5.3 to draw attention to additional graphical techniques that can assist in meeting the provisions of 10.1;

— Formulae B.4 and B.8 have been corrected to use $s_t^2$ instead of $w_t^2$;

— Formula B.16 has been corrected so that the term inside the square root is always non-negative;

— in Table C.2, the correction factor associated with $p = 2$ has been corrected to read 0,3994;

— additional literature references to the source of values in Table C.2 have been added to the Bibliography and referenced from Notes 1 and 2 of C.5.2.1;

— font styles (Italic or Roman) have been amended throughout for consistency in formulae.

© ISO 2022 – All rights reserved

# 0 Introduction

## 0.1 The purposes of proficiency testing

Proficiency testing involves the use of interlaboratory comparisons to determine the performance of participants (which may be laboratories, inspection bodies, or individuals) for specific tests or measurements, and to monitor their continuing performance. There are a number of typical purposes of proficiency testing, as described in the Introduction to ISO/IEC 17043. These include the evaluation of laboratory performance, the identification of problems in laboratories, establishing effectiveness and comparability of test or measurement methods, the provision of additional confidence to laboratory customers, validation of uncertainty claims, and the education of participating laboratories. The statistical design and analytical techniques applied shall be appropriate for the stated purpose(s).

## 0.2 Rationale for scoring in proficiency testing schemes

A variety of scoring strategies is available and in use for proficiency testing. Although the detailed calculations differ, most proficiency testing schemes compare the participant's deviation from an assigned value with a numerical criterion which is used to decide whether or not the deviation represents cause for concern. The strategies used for value assignment and for choosing a criterion for assessment of the participant deviations are therefore critical. In particular, it is important to consider whether the assigned value and criterion for assessing deviations should be independent of participant results, or should be derived from the results submitted. In this document, both strategies are provided for. However, attention is drawn to the discussion that will be found in Clauses 7 and 8 of the advantages and disadvantages of choosing assigned values or criteria for assessing deviations that are not derived from the participant results. It will be seen that in general, choosing assigned values and assessment criteria independently of participant results offers advantages. This is particularly the case for the criterion used to assess deviations from the assigned value – such as the standard deviation for proficiency assessment or an allowance for measurement error – for which a consistent choice based on suitability for a particular end use of the measurement results, is especially useful.

## 0.3 ISO 13528 and ISO/IEC 17043

This document provides support for the implementation of ISO/IEC 17043 particularly, on the requirements for the statistical design, validation of proficiency test items, review of results, and reporting summary statistics. ISO/IEC 17043:2010, Annex B, briefly describes the general statistical methods that are used in proficiency testing schemes. This document is intended to be complementary to ISO/IEC 17043, providing detailed guidance that is lacking in that document on particular statistical methods for proficiency testing.

The definition of proficiency testing in ISO/IEC 17043 is repeated in this document, with the notes that describe different types of proficiency testing and the range of designs that can be used. This document cannot specifically cover all purposes, designs, matrices and measurands. The techniques presented in this document are intended to be broadly applicable, especially for newly established proficiency testing schemes. It is expected that statistical techniques used for a particular proficiency testing scheme will evolve as the scheme matures; and the scores, evaluation criteria, and graphical techniques will be refined to better serve the specific needs of a target group of participants, accreditation bodies, and regulatory authorities.

This document incorporates published guidance for the proficiency testing of chemical analytical laboratories[32] but additionally includes a wider range of procedures to permit use with valid measurement methods and qualitative identifications. The revision of this document contains most of the statistical methods and guidance from the first edition, extended as necessary by the previously referenced documents and the extended scope of ISO/IEC 17043. ISO/IEC 17043 includes proficiency testing for individuals and inspection bodies, including ISO/IEC 17043:2010, Annex B, which includes considerations for qualitative results.

This document includes statistical techniques that are consistent with other International Standards, particularly those of TC69/SC6, notably the ISO 5725 series of standards on Accuracy: trueness and


© ISO 2022 – All rights reserved

precision. The techniques are also intended to reflect other International Standards, where appropriate, and are intended to be consistent with ISO/IEC Guide 98-3 (GUM) and ISO/IEC Guide 99 (VIM).

### 0.4 Statistical expertise

ISO/IEC 17043 requires that in order to be competent, a proficiency testing provider shall have access to statistical expertise and shall authorize specific personnel to conduct statistical analysis. Neither ISO/IEC 17043 nor this document can specify further what that necessary expertise is. For some applications an advanced degree in statistics is useful, but usually the needs for expertise can be met by individuals with technical expertise in other areas, who are familiar with basic statistical concepts and have experience or training in the common techniques applicable to the analysis of data from proficiency testing schemes. If an individual is responsible for statistical design and/or analysis, it is very important that this person has experience with interlaboratory comparisons, even if that person has an advanced degree in statistics. Conventional advanced statistical training often does not include exercises with interlaboratory comparisons, and the unique causes of measurement error that occur in proficiency testing can seem obscure. The guidance in this document cannot provide all the necessary expertise to consider all applications, and cannot replace the experience gained by working with interlaboratory comparisons.

### 0.5 Computer software

Computer software that is needed for statistical analysis of proficiency testing data can vary greatly, ranging from simple spread sheet arithmetic for small proficiency testing schemes using known reference values to sophisticated statistical software used for statistical methods reliant on iterative calculations or other advanced numerical methods. Most of the techniques in this document can be accomplished by conventional spread sheet applications, perhaps with customised routines for a particular proficiency testing scheme or analysis; some techniques will require computer applications that are freely available. In all cases, the users are expected to verify the validity and accuracy of their calculations, especially when special routines have been entered by the user. However, even when the techniques in this document are appropriate and correctly implemented by adequate computer applications, they cannot be applied without attention from an individual with technical and statistical expertise that is sufficient to understand the nature of the applications and the statistical assumptions, and to identify and investigate anomalies that can occur in any round of a proficiency testing scheme.

© ISO 2022 – All rights reserved

# Statistical methods for use in proficiency testing by interlaboratory comparison

## 1 Scope

This document provides detailed descriptions of statistical methods for proficiency testing providers to use to design proficiency testing schemes and to analyse the data obtained from those schemes. This document provides recommendations on the interpretation of proficiency testing data by participants in such proficiency testing schemes and by accreditation bodies.

The procedures in this document can be applied to demonstrate that the measurement results obtained by laboratories, inspection bodies, and individuals meet specified criteria for acceptable performance.

This document is applicable to proficiency testing where the results reported are either quantitative measurements or qualitative observations on test items.

NOTE The procedures in this document can also be applied for the assessment of expert opinion where the opinions or judgments are reported in a form which can be compared objectively with an independent reference value or a consensus statistic. For example, when classifying proficiency test items into known categories by inspection - or in determining by inspection whether proficiency test items arise, or do not arise, from the same original source - and the classification results are compared objectively, the provisions of this document that relate to nominal (qualitative) properties can be applied.

## 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 3534-2, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*

ISO/IEC 17043, *Conformity assessment — General requirements for proficiency testing*

ISO Guide 30, *Reference materials — Selected terms and definitions*

ISO/IEC Guide 99, *International vocabulary of metrology — Basic and general concepts and associated terms (VIM)*

## 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 3534-2, ISO 5725-1, ISO/IEC 17043, ISO/IEC Guide 99, ISO Guide 30, and the following apply. In the case of differences between these references on the use of terms, definitions in ISO 3534-1 ISO 3534-2 apply. Mathematical symbols are listed in Annex A.

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

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

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

© ISO 2022 – All rights reserved

**3.1**
**interlaboratory comparison**
organization, performance and evaluation of measurements or tests on the same or similar items by two or more laboratories in accordance with predetermined conditions

**3.2**
**proficiency testing**
evaluation of participant performance against pre-established criteria by means of *interlaboratory comparisons* (3.1)

Note 1 to entry: For the purposes of this document, the term "proficiency testing" is taken in its widest sense and includes, but is not limited to:

— quantitative scheme — where the objective is to quantify one or more measurands for each proficiency test item;

— qualitative scheme — where the objective is to identify or describe one or more qualitative characteristics of the proficiency test item;

— sequential scheme — where one or more proficiency test items are distributed sequentially for testing or measurement and returned to the proficiency testing provider at intervals;

— simultaneous scheme — where proficiency test items are distributed for concurrent testing or measurement within a defined time period;

— single occasion exercise — where proficiency test items are provided on a single occasion;

— continuous scheme — where proficiency test items are provided at regular intervals;

— sampling — where samples are taken for subsequent analysis and the purpose of the proficiency testing scheme includes evaluation of the execution of sampling; and

— data interpretation — where sets of data or other information are furnished and the information is processed to provide an interpretation (or other outcome).

**3.3**
**assigned value**
value attributed to a particular property of a proficiency test item

**3.4**
**standard deviation for proficiency assessment**
measure of dispersion used in the evaluation of results of *proficiency testing* (3.2)

Note 1 to entry: This can be interpreted as the population standard deviation of results from a hypothetical population of laboratories performing exactly in accordance with requirements.

Note 2 to entry: The standard deviation for proficiency assessment applies only to ratio and interval scale results.

Note 3 to entry: Not all proficiency testing schemes evaluate performance based on the dispersion of results.

[SOURCE: ISO/IEC 17043:2010, modified — In the definition "based on the available information" has been deleted. Note 1 to the entry has been added, and Notes 2 and 3 have been slightly edited.]

**3.5**
**measurement error**
measured quantity value minus a reference quantity value

[SOURCE: ISO/IEC Guide 99:2007, modified — Notes have been deleted.]


© ISO 2022 – All rights reserved

**3.6**
**maximum permissible error**
extreme value of *measurement error* (3.5), with respect to a known reference quantity value, permitted by specifications or regulations for a given measurement, measuring instrument, or measuring system

[SOURCE: ISO/IEC Guide 99:2007, modified — Notes have been deleted.]

**3.7**
***z* score**
standardized measure of performance, calculated using the participant result, *assigned value* (3.3) and the *standard deviation for proficiency assessment* (3.4)

Note 1 to entry: A common variation on the *z* score, sometimes denoted z' (commonly pronounced *z*-prime), is formed by combining the uncertainty of the assigned value with the standard deviation for proficiency assessment before calculating the *z* score.

**3.8**
**zeta score**
standardized measure of performance, calculated using the participant result, *assigned value* (3.3) and the combined standard uncertainties for the result and the *assigned value* (3.3)

**3.9**
**proportion of allowed limit score**
standardized measure of performance, calculated using the participant result, *assigned value* (3.3) and the criterion for *measurement error* (3.5) in a proficiency test

Note 1 to entry: For single results, performance can be expressed as the deviation from the assigned value (*D* or *D %*).

**3.10**
**action signal**
indication of a need for action arising from a proficiency test result

EXAMPLE A *z* score in excess of 2 is conventionally taken as an indication of a need to investigate possible causes; a *z* score of 3 or greater is conventionally taken as an action signal indicating a need for corrective action.

**3.11**
**consensus value**
value derived from a collection of results in an *interlaboratory comparison* (3.1)

Note 1 to entry: The phrase 'consensus value' is typically used to describe estimates of location and dispersion derived from participant results in a round of a proficiency testing scheme, but may also be used to refer to values derived from results of a specified subset of such results or, for example, from a number of expert laboratories.

**3.12**
**outlier**
member of a set of values which is inconsistent with other members of that set

Note 1 to entry: An outlier can arise by chance from the expected population, originate from a different population, or be the result of an incorrect recording or other blunder.

Note 2 to entry: Many proficiency testing schemes use the term outlier to designate a result that generates an action signal. This is not the intended use of the term. While outliers will usually generate action signals, it is possible to have action signals from results that are not outliers.

[SOURCE: ISO 5725-1:1994, modified — The Notes to the entry have been added.]

**3.13**
**participant**
laboratory, organization, or individual that receives proficiency test items and submits results for review by the *proficiency testing* (3.2) provider

© ISO 2022 – All rights reserved

**3.14**
**proficiency test item**
sample, product, artefact, reference material, piece of equipment, measurement standard, data set or other information used to assess *participant* (3.13) performance in *proficiency testing* (3.2)

Note 1 to entry: In most instances, proficiency test items meet the ISO Guide 30 definition of *"reference material"* (3.17).

**3.15**
**proficiency testing provider**
organization which takes responsibility for all tasks in the development and operation of a *proficiency testing* (3.2) scheme

**3.16**
**proficiency testing scheme**
*proficiency testing* (3.2) designed and operated in one or more rounds for a specified area of testing, measurement, calibration or inspection

Note 1 to entry: A proficiency testing scheme might cover a particular type of test, calibration, inspection or a number of tests, calibrations or inspections on proficiency test items.

**3.17**
**reference material**
**RM**
material, sufficiently homogeneous and stable with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process

Note 1 to entry: RM is a generic term.

Note 2 to entry: Properties can be quantitative or qualitative, e.g. identity of substances or species.

Note 3 to entry: Uses may include the calibration of a measuring system, assessment of a measurement procedure, assigning values to other materials, and quality control.

[SOURCE: ISO Guide 30:2015, modified —Note 4 has been deleted.]

**3.18**
**certified reference material**
**CRM**
*reference material (RM)* (3.17) characterized by a metrologically valid procedure for one or more specified properties, accompanied by an RM certificate that provides the value of the specified property, its associated uncertainty, and a statement of metrological traceability

Note 1 to entry: The concept of value includes a nominal property or a qualitative attribute such as identity or sequence. Uncertainties for such attributes may be expressed as probabilities or levels of confidence.

[SOURCE: ISO Guide 30:2015, modified —Notes 2, 3 and 4 have been deleted.]

# 4 General principles

## 4.1 General requirements for statistical methods

**4.1.1** The statistical methods used shall be fit for purpose and statistically valid. Any statistical assumptions on which the methods or design are based shall be stated in the design or in a written description of the proficiency testing scheme, and these assumptions shall be demonstrated to be reasonable.

NOTE A statistically valid method has a sound theoretical basis, has known performance under the expected conditions of use and relies on assumptions or conditions which can be shown to apply to the data sufficiently well for the purpose at hand.


© ISO 2022 – All rights reserved