INTERNATIONAL STANDARD

ISO 9241-304

First edition
2008-11-15

Ergonomics of human-system interaction —

Part 304: User performance test methods for electronic visual displays

Ergonomie de l'interaction homme-système

Partie 304: Méthodes d'essai de la performance de l'utilisateur pour écrans de visualisation électroniques

Reference number
ISO 9241-304:2008(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web 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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 9241-304 was prepared by Technical Committee ISO/TC 159, *Ergonomics*, Subcommittee SC 4, *Ergonomics of human-system interaction*.

This first edition of ISO 9241-304, together with ISO 9241-302:2008, ISO 9241-303:2008, ISO 9241-305:2008 and ISO 9241-307:2008, cancels and replaces ISO 9241-3:1992, of which it constitutes a technical revision. It also incorporates the Amendment ISO 9241-3:1992/Amd.1:2000, replacing that Amendment's test method with the one specified in Clause 8.

ISO 9241 consists of the following parts, under the general title *Ergonomic requirements for office work with visual display terminals (VDTs)*:

— *Part 1: General introduction*

— *Part 2: Guidance on task requirements*

— *Part 4: Keyboard requirements*

— *Part 5: Workstation layout and postural requirements*

— *Part 6: Guidance on the work environment*

— *Part 9: Requirements for non-keyboard input devices*

— *Part 11: Guidance on usability*

— *Part 12: Presentation of information*

— *Part 13: User guidance*

— *Part 14: Menu dialogues*

— *Part 15: Command dialogues*

— *Part 16: Direct manipulation dialogues*

— *Part 17: Form filling dialogues*

ISO 9241 also consists of the following parts, under the general title *Ergonomics of human-system interaction*:

— *Part 20: Accessibility guidelines for information/communication technology (ICT) equipment and services*

— *Part 110: Dialogue principles*

— *Part 151: Guidance on World Wide Web user interfaces*

— *Part 171: Guidance on software accessibility*

— *Part 300: Introduction to electronic visual display requirements*

— *Part 302: Terminology for electronic visual displays*

— *Part 303: Requirements for electronic visual displays*

— *Part 304: User performance test methods for electronic visual displays*

— *Part 305: Optical laboratory test methods for electronic visual displays*

— *Part 306: Field assessment methods for electronic visual displays*

— *Part 307: Analysis and compliance test methods for electronic visual displays*

— *Part 308: Surface-conduction electron-emitter displays (SED)* [Technical Report]

— *Part 309: Organic light-emitting diode (OLED) displays* [Technical Report]

— *Part 400: Principles and requirements for physical input devices*

— *Part 410: Design criteria for physical input devices*

— *Part 920: Guidance on tactile and haptic interactions*

For the other parts under preparation, see Annex A.

Introduction

ISO 9241 was originally developed as a seventeen-part International Standard on the ergonomics requirements for office work with visual display terminals. As part of the standards review process, a major restructuring of ISO 9241 was agreed to broaden its scope, to incorporate other relevant standards and to make it more usable. The general title of the revised ISO 9241, “Ergonomics of human-system interaction”, reflects these changes and aligns the standard with the overall title and scope of Technical Committee ISO/TC 159, Subcommittee SC 4. The revised multipart standard is structured as series of standards numbered in the “hundreds”: the 100 series deals with software interfaces, the 200 series with human centred design, the 300 series with visual displays, the 400 series with physical input devices, and so on.

See Annex A for an overview of the entire ISO 9241 series.

ISO 9241-3:1992, Annex C, offered users a provisional alternative method for testing the visual quality of a display, intended for novel display technologies for which no optical test method was available. The Amendment ISO 9241-3:1992/Amd.1:2000 replaced this test method and made the previously informative Annex C normative. ISO 9241-7:1998, ISO 9241-8:1997 and ISO 13406-2:2001 (all three of which have since been cancelled and replaced by other parts of the ISO 9241 “300” subseries) referred to that Amendment as providing an alternative user performance test method.

This part of ISO 9421 not only incorporates the Amendment, but extends its basis to provide guidance on the general process of assessing the visual ergonomics of displays in a specific context of use by means of a user performance test method. The test method specified in this part of ISO 9241 is applicable only to user tasks involving the handling and processing of text. However, it is expected that test procedures will also be developed for using maps and for handling and interpreting photographs and moving images, with these then being incorporated into a future edition.

The structure of this part of ISO 9241 is an exception in the ISO 9241 “300” subseries in that it establishes the conformance of a visual display used for text rendition according to its own user performance test method, instead of by means of a compliance route given in ISO 9241-307 (in which no compliance route relevant to this part of ISO 9241 is provided).

Ergonomics of human-system interaction —

Part 304:
User performance test methods for electronic visual displays

1 Scope

This part of ISO 9241 provides guidance for assessing the visual ergonomics of display technologies with user performance test methods (as opposed to the optical test methods given in ISO 9241-305). Its use will help to ensure that, for a given context of use, a display meets minimum visual ergonomics requirements. It covers only visual attributes and does not address the ergonomics or usability of the whole product that houses a visual display.

The general principles laid down by this part of ISO 9241 apply to any colour or monochrome visual display attached to a system with which human beings interact. This includes, but is not limited to, visual displays used with desktop and portable computers, those used on mobile devices such as mobile telephones, digital cameras and personal digital assistants, and status displays used on consumer electronics equipment such as printers, in-car navigation systems and microwave ovens. It extends the basic idea of the visual performance and comfort test specified in ISO 9241-3:1992/Amd.1:2000 to the use of the performance and judgment of the display end users themselves for evaluating the quality of a display, and includes a more diverse range of technologies, users, tasks and environments.

Because of this diversity, it is not feasible for this part of ISO 9241 to stipulate a single, generic test method that can be used with all display technologies. Instead, the basic principles for generating a test method are given. This method will be valid for evaluating specific displays in specific contexts of use: the method generated according to Clause 8 is applicable only to tasks involving the handling and processing of text. No other examples are given. An essential property of the process is that it permits the verification of the usability of a visual display for a representative task, performed by representative users, taking their performance and judgment as measured quality values. It does not, however, permit the measurement of specific perceptual attributes such as luminance contrast or display flicker in isolation.

The main users of this part of ISO 9241 will be those who procure displays or who need to measure display performance during product development. Its application assumes a background in behavioural science.

2 Normative references

The following referenced documents are indispensable for the application 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 9241-5, *Ergonomic requirements for office work with visual display terminals (VDTs) — Part 5: Workstation layout and postural requirements*

ISO 9241-6, *Ergonomic requirements for office work with visual display terminals (VDTs) — Part 6: Guidance on the work environment*

ISO 9241-302, *Ergonomics of human-system interaction — Part 302: Terminology for electronic visual displays*

ISO 9241-303:2008, *Ergonomics of human-system interaction — Part 303: Requirements for electronic visual displays*

ISO/IEC 8859 (all parts), *Information technology — 8-bit single-byte coded graphic character sets*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 9241-302 apply.

4 Guiding principles

The guiding principles of this part of ISO 9241 are that visual displays should help people carry out their tasks effectively and efficiently, and that displays should be satisfying to use and not in any way be harmful to their users' health.

Formal optical test methods as specified in ISO 9241-305 might not be available to support the procurement of newer display technologies. In such cases, requiring manufacturers to demonstrate evidence of the usability of their displays provides the most effective route for ensuring good ergonomics quality. This is the approach taken by this part of ISO 9241.

It sets out four steps for generating test methods that can be used to measure the ergonomics visual quality of visual displays:

a) specify the visual ergonomic test objectives (see Clause 6);

b) define the test procedure (see Clause 7);

c) carry out the test (see Clause 8);

d) analyse the data (see Clause 8).

5 Conformance

Whereas ISO 9241-303 and ISO 9241-305 refer to the compliance routes defined in ISO 9241-307 to establish the conformance of a visual display, this part of ISO 9241 itself specifies a test method for establishing such conformance.

If the test display is compared to a benchmark display and the test procedure is based on either alphanumeric or non-alphanumeric text, conformance is achieved when both

— the search speed for the test display is not statistically significantly lower than the search speed obtained with the benchmark display, and

— the perceived quality, as measured by its visual comfort rating, of the test display is not statistically significantly lower than that of the benchmark display.

The procedure used for determining search speed and perceived quality shall be in accordance with Clause 8.

6 Specifying the visual ergonomics test objectives

6.1 General

Visual ergonomics can be measured like any other engineering attribute. Although the data from user performance tests are derived from objective and subjective measures of human performance, this does not mean that the data are simply personal opinions. A good test design will generate data that are objective and unbiased. Useful information on many practical aspects of test design in general is given in ISO 20282; ISO 20282-1, in particular, provides valuable background information in this area.

Testing only makes sense if the test results are compared with criteria that define a display as acceptable or unacceptable. The aim of this step is to define those criteria for the display to be tested.

6.2 Criterion description

The criterion description defines the context of the measurements and the performance characteristics that will be measured. In most cases — when, for instance, a novel display technology is used in a visual display that helps perform an existing task (such as word processing in an office) — the visual quality of the test display is assessed against that of a benchmark display known to meet or exceed the requirements of ISO 9241-303, using a measuring method according to ISO 9241-305 and a conformance method according to ISO 9241-307.

EXAMPLE A test of a display that will be used for in-car navigation might use as the criterion: “Ease of reading information from the display when it is used by experienced drivers in bright ambient lighting”.

6.3 Measuring method

The measuring method describes how the criterion will be measured, i.e. the scale that will be used for the measurement and how the values will be derived.

As an example, in ISO 9241-11, three separate measures are taken:

— effectiveness (the accuracy and completeness with which customers achieve specified goals);

— efficiency (the accuracy and completeness of goals achieved in relation to resources);

— satisfaction (freedom from discomfort, and positive attitudes towards the use of the visual display).

It needs be realized that these three measures from ISO 9241-11 are context-dependent; this means that, for example, the effectiveness of a mobile phone display could be low — even very low compared to that of a desktop display — yet entirely satisfactory in the context of mobile phone use.

EXAMPLE A test of a display used on a mobile phone measures the accuracy with which a participant can distinguish different colours (effectiveness), the speed with which a participant can read text on the display (efficiency) as well as the participant's overall attitude towards the display's image quality (satisfaction).

6.4 Performance criteria

If the test display is compared to a benchmark display, the performance criterion is normally that the test display have at least the same visual quality as the benchmark one. But in other cases, making a decision on the value that is acceptable can require some market analysis. Useful questions to ask at this stage include the following.

— Is there an earlier version of the tested display that is relevant in this context? If so, how is its visual quality rated?

— How do competitors displays perform?

These values provide the engineer with a lower limit to place on the performance of the display. Human factors specialists recommend that the response range be considered as a continuum ranging from “Unacceptable”, through a “Minimum” range into a “Target” range, and finally into an “Exceeds” range, as follows.

a) ***Unacceptable***

If the display performs within this range, it cannot be released.

b) ***Minimum***

If the display performs within this range, it is barely acceptable. Management must weigh the benefits of releasing a barely adequate display now, versus waiting for the usability defects to be fixed.

c) ***Target***

If the display performs within this range, it can be released. This is the performance range thought to be necessary to succeed.

d) ***Exceeds***

If the display performs within this range, it could be that the development team have put too much effort into the design of the display and/or developed a product surpassing requirements.

This approach makes it unlikely that the development team will either under- or over-engineer the display.

7 Defining the test procedure

7.1 General

The test procedure shall be oriented towards a task carried out with the help of a visual display: a user performance test method as defined in this part of ISO 9241 relies on a user behaving as when performing a typical task of this kind.

7.2 Alphanumeric and non-alphanumeric text

The visual performance and comfort test specified in Clause 8 may be used to establish the conformance of a visual display to a certain quality, according to user performance, applicable to tasks involving the handling and processing of text. Its test procedure is suitable for such tasks, typical of an office environment. Test methods for other types of tasks, with their appropriate devices, remain to be developed.

NOTE 1 So far, no test procedure for using maps has been developed.

NOTE 2 So far, no test procedure for handling and interpreting photographs has been developed.

NOTE 3 So far, no test procedure for handling and interpreting moving images has been developed.

8 Visual performance and comfort test — Carrying out the test and analysing the data

8.1 General

This clause specifies a method that can be used to test the visual quality of electronic visual displays where the entire set of physical requirements (viz. ISO 9241-303) cannot be applied — for example, in the case of novel display technologies such as new types of flat panel display (e.g. electroluminescent or field emission displays). The test consists of a letter search task and an assessment of visual comfort.

The context of use is *text processing*, such as that which takes place in office work. The combination of test results will be referred to as the *visual quality* of the display.

The various stages of the test, such as recruiting participants, running a pilot test and administering the final test are described (see also ISO/TS 20282-2), as is the analysis of the test data, including application of statistics.

8.2 Purpose

The method tests displays that cannot otherwise be tested for conformance with ISO 9241. It is not an alternative test method in the sense that a display manufacturer has the option of choosing either the physical requirements described in ISO 9241-303 or this method. Instead, the test method provides a testing route for displays that cannot be tested according to the requirements specified in ISO 9241-303 because the display under test uses a new technology that makes metrological measurements such as those specified in ISO 9241-305 difficult or impossible.

8.3 Overview

8.3.1 General

This test procedure measures the effectiveness of the transfer of visual information in terms of participants' search performance for targets embedded in alphanumerics on a candidate display versus those same participants' performance for such a task on a benchmark display. Effectiveness in this context means that the user is able to detect and recognize the visual targets accurately, quickly and without visual discomfort. If the display passes this particular visual search and rating performance test, it can be assumed capable of other presentations of information such as, but not limited to, non-alphanumeric languages.

NOTE For a general background on this type of test methodology, see ISO/IEC 25062.

The dependent variables of the test are the search speed achieved by the test participants in a visual letter search task and subjective ratings of visual comfort using a category scale. Testing takes place in a simulated office environment, with test participants representative of the anticipated user population.

The method, i.e. a letter search task applying pseudo-text in combination with scaling of experienced visual comfort, was first developed and tested by researchers of the centre for research on user-system interaction (IPO) [12], [13].

The visual quality of a display, referred to as a test display, is assessed against a benchmark display known to meet or exceed the requirements of ISO 9241-303. Both the speed in the visual search task and the subjective ratings must meet certain minimum requirements for the test display to pass. Sequential statistics, or an equally robust equivalent statistical procedure, are used to determine if the participant's performance on the test display exceeds, is equal to, or falls short of, performance on the benchmark display.

8.3.2 Avoidance of bias

All tests are open to bias, and this is especially true in the area of psychological testing. The assessment shall therefore be carried out under the supervision of those qualified to carry out such testing, having the necessary education and at least one year of experience. Rules governing the ethical conduct of human experimental testing shall be followed. Examples of such rules can be found in References [14] and [15].

The test administrator should ensure that all potential sources of error are minimized or controlled. The following list describes some potential sources of bias and error; it is not intended to be exhaustive.

— Selection of test participants: for example, avoid selecting particular age groups.

— Configuration of displays: during the test, the benchmark display shall meet all requirements of ISO 9241-303.

— Environmental conditions: lighting and other conditions shall be equal for both displays, in order to avoid detrimental conditions for one of them.

— Instructions to the test participants: these should be impartial.

8.4 Test participants

Test participants should be a sample representing the anticipated user population. The following assumes this population to have unimpaired vision; for people with visual impairments using visual displays, other criteria for required image quality apply, and other tests would need to be designed. All test participants shall have visual acuity that is normal, or corrected to normal, at the design viewing distance, and they shall be without any obvious physical or physiological conditions that could influence either their search performance or the quality of the images that they perceive.

8.5 The displays

The test display shall be a production or full-feature pre-production unit. It shall incorporate any treatments that would be in the production unit, for example, surface treatments such as anti-glare and anti-reflection filters and treatments. The benchmark display shall be supplied or nominated by the supplier of the test display and shall meet or exceed all requirements of ISO 9241-303; it shall have approximately the same size as the test display.

The displays may be labelled for identification purposes: for example, as “Display 1” and “Display 2”. Under these conditions, the test participants should not be informed as to which is the test and which the benchmark display, with half the number of test participants first having “Display 1”, the test display, and the other half first having “Display 2”, the benchmark display.

8.6 Test setup

8.6.1 General

The test shall be conducted in an area that is free from distractions and external interference that could influence the test results. The ambient conditions shall fall within the range defined in ISO 9241-6. These conditions shall be comfortable and shall not be subject to significant variation during the test, either within a test participant's session or between test participants.

8.6.2 Test environment

The thermal environment, the background noise level, the ambient lighting and the reflectance of work surfaces shall meet the minimum requirements set out in ISO 9241-5 and ISO 9241-6. The ambient illumination shall be designed so as to minimize glare and specular reflections. Constant lighting conditions shall be maintained both within a participant's session and between test participants. The test participants shall be light-adapted by being placed in the test room for 10 min prior to the test.

NOTE This period can be used by the experimenter to instruct each test participant about the test.

8.6.3 Test workstation

The display and associated equipment (e.g. the keyboard) shall be supported by a work surface in accordance with ISO 9241-5.

For both the benchmark and test displays, the viewing distance shall be set according to the design viewing distance. This distance should be constrained by a head-and-chin rest, the height of which is adjustable. The individual height adjustments for the test participants should be such that for both the test display and the benchmark display, the position of the eyes with respect to the display is equal for all test participants. The position of the test participant's eyes shall comply with the line-of-sight angle requirements given in ISO 9241-303:2008, 5.1.4.

The brightness and contrast settings of the benchmark display shall be specified by the manufacturer who nominates the display: at these settings it shall meet or exceed all respective requirements given in ISO 9241-303.

NOTE A measurement procedure for specifying brightness and contrast is given in ISO 9241-305.

At the option of the manufacturer, the brightness and contrast of the test display should be either

a) fixed at settings specified by the manufacturer, or

b) adjustable by test participants to their personal optimum settings.

Both displays shall be allowed to warm-up for at least 20 min prior to testing.

The test participant shall be seated in accordance with ISO 9241-5.

8.6.4 Test material

The test material shall be pseudo-text generated from a character set associated with an 8-bit single-byte coded graphic character set as specified ISO/IEC 8859, which describes a collection of character sets for various languages. If a system cannot display text in an alphabet familiar to the users, text should be displayed by double-byte coded characters (e.g. Asian characters). In this case, the language used shall be specified in the compliance statement. Each test will use a specified character subset (e.g. “A” to “Z”, “a” to “z”, and “0” to “9”). The same subset shall be used for both displays.

Pseudo-text shall be generated from the character set according to the following constraints.

— Pseudo-text shall consist of blocks of random strings of characters separated by spaces.

— The texts, on both test and benchmark displays, shall consist of a constant number of lines and a constant number of characters per line (including space characters).

— The number of characters per line shall be chosen so that the line length (in centimetres) is less than 25 times the line-to-line distance (i.e. the height of the display area divided by the maximum number of lines). However, a line should contain at least 30 characters (including embedded spaces). The total number of characters in a pseudo-text shall be between 400 and 600, embedded spaces included. The pseudo-text blocks shall be sized such that, if five blocks could be displayed at once (one in each corner and one in the middle), they would have minimum overlap while maximizing coverage of the display area.

— Each test participant shall be instructed to count the occurrences of a single target character over the entire test (e.g. test participant X is instructed to search for “A”s during the entire test, test participant Y is instructed to search for “R”s, etc.).

— The number of targets shall be 2 % to 3 % of the total number of characters in the text, including embedded spaces.

— The position of the targets shall be randomly chosen with the restriction that a line shall not start or end with the target character.

— The text shall contain a constant number of spaces. The space fraction shall be 15 % (i.e. the number of spaces relative to the total number of characters, including embedded spaces).

NOTE Although the average word length does vary over different languages, pseudo-texts with 15 % space fraction do resemble normal texts with respect to their string length distributions.

The position of the spaces shall be randomly chosen with the following restriction:

a) a line shall neither start nor end with a space character (all spaces are embedded);

b) a space character shall not be adjacent to another space character (strings are separated by single spaces);

c) the minimum string length shall be two characters.

8.6.5 Test procedure

Display pseudo-text as a block of characters in one of five screen locations. The test participant's task is to scan the text and identify each instance of the target character.

Place the blocks of pseudo-text in the upper left, upper right, lower left, lower right and centre of the screen. Locate the centre block so that the middle character of the block is approximately in the centre of the active area of the screen. Place text in each of the four corners so that it abuts the extreme corners of the screen.

Inform the test participants that the objective of the test is to evaluate the quality of the image on the display. If, for the purposes of the experiment, the manufacturer of the test display has decided that the brightness and contrast may be adjusted by test participants, give the test participants the opportunity to adjust the test display to their preferred settings. Set the brightness and contrast settings of the benchmark display in accordance with the manufacturer's instructions. This shall not be adjusted by the test participant.

Manufacturers should be aware that, if the user is allowed to adjust the display, this can give the user an indication of the display under test and therefore could affect the results of the test. This may be prevented by asking the user to adjust the settings before the test and then performing the test with the controls hidden from view.

Present the five test blocks at the five locations in random order. Instruct each test participant to scan the pseudo-text from the top to the bottom line and indicate each occurrence of the target character. In order to overcome the problem of initial learning effects, train the test participants before the main experiment by performing the task for at least 10 pseudo-texts (i.e. 10 trials). Residual learning shall be controlled by counter-balancing the stimulus order within the main experiment. These practice trials shall use pseudo-text placed in any of the five possible screen locations. Practice trials shall be presented on both test and benchmark displays.

Continue practice trials until the test participant's performance on any one block of pseudo-text is error-free. Do not use data collected from the practice trials to evaluate the quality of the display.

For the experimental trials, measure the time taken for the test participant to identify each instance of the target character in each block of pseudo-text and the number of errors made by the test participant (see 8.8).

Allow the test participant a rest break of up to 1 min between trials, with a minimum break of 10 s.

Instruct test participants to respond by pressing predefined keys or buttons to: *initiate* a trial; *count* spotted targets; and *stop* a trial.

A keyboard or any other appropriate input device may be used for this purpose. If the keyboard is used, the “ENTER” key should be defined to initiate/stop a trial, and the space bar should be defined to register spotted targets.

Register the interval between initiation and stopping of a trial as the search time for this trial.

Instruct test participants to minimize errors as far as possible and yet work as quickly as possible. They shall be instructed to minimize their error rate, regardless of the visual quality of the display under test; for example, if the display has deteriorated in comparison to a previous one. It is very important to give the proper instruction to the test participants in this respect, who then generally are well able to keep their error rate constant and low.

Half of the test participants shall use the benchmark display first, and the other half shall use the test display first.

On completion of the visual search task with a display, ask the test participants to rate the visual quality of that display on a nine-point numerical scale (shown below), with 1 being “Poor” and 9 being “Excellent”. After completion of the trials with the test display or the benchmark display, ask the test participants to assess the perceptual quality of that display with respect to its visual comfort.

1	2	3	4	5	6	7	8	9
Poor			Fair			Excellent		

The following written instructions shall be given to the test participants to explain how responses are to be made.

> “We would like you to indicate how you judge the display you have just used with respect to its visual comfort. Please circle the number corresponding to your judgement.”

A sample set of instructions for test participants is given in 8.6.7.

8.6.6 Task conditions

Display attributes (character size, resolution, visual angle, fonts, etc.) of the test display and the benchmark display shall be specified by the manufacturer who nominates the display. These attributes shall be stated in the compliance statement.

— The same font shall be used on both the test and the benchmark display. This font shall be a fixed-width font which complies with the requirements for size, shape and spacing of characters given in ISO 9241-303.

— For each test participant, a fixed target character shall be used over the whole experiment.

— A target character shall have average, i.e. neither too low nor too high, discriminability with the other characters used — for example, do not use “O”, “0” or “Q”). This test method is not intended to evaluate font design.

— The number of target occurrences shall be variable over different pseudo-texts.

— The total number of targets over all trials shall be constant for each display. The test participants shall not be informed of these totals.

— The number of different pseudo-texts per test participant shall be large enough to prevent memorizing effects. An appropriate number is 20 (or less if the number of trials per test participant is less).

— The pseudo-texts shall be presented counterbalanced over all conditions (displays) and/or test participants.

— The test participants should scan the text line-by-line, each line either from left to right or from right to left, according to the direction of reading that they apply in their native language.

— Search time shall begin immediately after the pseudo-text is presented on the display. Search time shall end when the test participant indicates completion of the page of pseudo-text.

— The test participants shall use a button (or key on keyboard) each time a target is spotted. The number of counted targets shall be registered as a check of the test participant's concentration. The performance measurement shall be neglected from statistical treatment if the recorded number of targets differ by $\geqslant$ 10 % from the actual number of targets in the block.

— The test participants shall use another button (or key) to start/stop time registration.

8.6.7 Instructions to test participants

These are sample instructions that should be modified for the particular testing situation. They assume that keyboard input is to be used in a country in which the direction of reading the native language is from left to right and will need to be modified if another direction of reading or a non-keyboard input device is used.

The instructions shall be presented to the participants on paper; an example of how they might read follows.

> “Thank you for taking part in this test. The aim of this test is to evaluate character legibility. Please remember that we are testing the display(s) and not you!
>
> “You will be presented with a series of screens similar to the example below. Your task is to find each capital letter “A”. You should read the text from the top left to the bottom right, as if you are reading a normal page of text. When you are ready to start a trial, press the *ENTER* key on the keyboard. You start your search immediately after a pseudo-text appears on one of five locations on the display (top-left, top-right, bottom-left, bottom-right or in the centre). Whenever you see a capital letter “A”, press the *space bar* on the keyboard. After you have finished reading the entire text, press the *ENTER* key again. Please work through the screens as accurately and as quickly as possible. The number of targets in each screen varies, so please pay careful attention to properly reading, searching and indicating the presence of the target letter in each screen in the series as accurately and also as quickly as possible. This means that if the visual quality of the display under test has deteriorated in comparison to a previous one you have to work slower, but if it has improved in comparison to a previous one you have to work faster. If you have any questions, please ask the test administrator now.

```
WhwNdzo zltpVY 1CCAe kDw he t3
TkW3rm8U ya BpE O2B L8Y A5 She
PQtb 90DViRCDG 1H pSM yEqZz 6F
jyA3 sATQesa ANUU VLH Ou1p2JBE
vbR l1Y5rVr SA9mr DmPETLV 2uO2
7phnFd2oyT 83ee zKo8h KyiTJgAL
vXMu 6Kugm 3ElkxsOWhCK1FTMA T6
LuGF5 ad HsicT H0jkHv ssAq U8Q
8dW rmrtfGqh HCsnGdYIMQEITS fo
o1 XVw6 2VogMFo6 PH uJD3c DXj8
yW 5LN 6Bv0 fGPhdZ Cn x9gUiaH3
fySFoauaxj UeK bKQz 2uZa MmnCN
4t HT3OFuMUSo piq1uUh8tdRbK1Tn
Ez 33Q 6w fvVR 7B gyz Ns5 5Ami
7T5k 6bc2 ZHl fJmDO GwJ9 ECKYm
Xob3m t9 SU ZR e1 3lFg 1wc j4w
nToPDF RCUb nyMHs rMI0oizFL8dx
a2Z sD AK5R1 Q8jiI wBeeA L2Rz0
```

”

8.7 Dependent measures

8.7.1 General

Two dependent measures shall be recorded from the experimental trials for each test participant. Data from the practice trials shall not be used in the following analysis.

The dependent measures shall be the following:

a) the average search speed obtained from trials with error rates < 10 %;

b) the subjective ratings of visual comfort.

Error rate, E, is defined as:

$$E = \frac{|T_0 - T_C|}{T_0} \times 100\ \%$$

where

T_0 is the total number of target characters in the page of pseudo-text shown to the test participant;

T_C is the total number of target characters counted by the test participant.

The performance measurement shall be neglected from statistical treatment if the number of missed or extra targets is too large (one missed or extra target is accepted in a text with 10 targets).

8.7.2 Average search speed

From the registered search time, T_i, corresponding to the valid trials ($E < 10$ %), the performance measure of a test participant, the *average search speed*, v_s, measured in characters/s, is calculated by:

$$v_s = n_t \cdot n_c \cdot \left[\sum_{i=1}^{n_t} T_i \right]^{-1}$$

where

n_t is the number of valid trials for that test participant;

n_c is the total number of characters in a pseudo-text (including embedded spaces).

NOTE The v_s values for the test and benchmark displays can be analysed by applying a sequential testing procedure for successive test participants (see 8.8).

8.7.3 Subjective ratings

Each test participant shall give, for both the test and the benchmark displays, a subjective rating of visual comfort on a nine-point scale.

NOTE These ratings can be analysed using the sequential testing procedure for successive test participants described in 8.8.

8.8 Statistical treatment of results

8.8.1 General

If sequential analysis is used for conducting conformance testing it can reduce the number of participants required to achieve a statistically reliable test of the null hypothesis.

NOTE The main feature of sequential analysis is that the sample size is not determined in advance; instead, the validity of the null hypothesis is tested after each set of results has been collected.

Other statistical procedures and analysis, for example, a t-test, may be carried out as long as an adequate sample size is used.

If the below sequential analysis procedure is not used, it shall be ensured in the test and statistical analysis that the Type 2 error rate, β, is smaller than 0,05 for a standard deviation, D, of 0,5, and that the criterion, α, manufacturer's risk, shall be 0,05 (see Table 1).

Statistical treatment of the results involves comparing the dependant measures for the test display against a benchmark. Since no statistical tests can prove that two products are the same, this test is used to decide whether performance for the test product is significantly worse or better than the benchmark. If the test product is not significantly worse than the benchmark, the test product is considered to conform to the standard.

Hence, the null hypothesis, H_0, is that the scores of the test display are equal to or better than those for the benchmark display. The alternative hypothesis, H_1, is that the scores for the test display are significantly worse than those for the benchmark display.

8.8.2 General theory

Statistical decisions are prone to two kinds of error. The first type of error (Type 1) occurs when the null hypothesis is falsely rejected; the second type (Type 2) when the null hypothesis is falsely not rejected. These two risks are usually symbolized by α and β (see Table 1).

Table 1 — Types of decision that can be made using a statistical test

	Decision after testing	
	Test display accepted	**Test display rejected**
Test display at least as good as benchmark display	Correct decision	Error Type 1: manufacturer's risk, α
Test display worse than benchmark display	Error Type 2: user's risk, β	Correct decision

In non-sequential testing, the sample size in an experiment shall be fixed in advance by using the following formula adapted from Reference [16]:

$$N = \frac{2\left(\mu_\alpha + \mu_\beta\right)^2}{D^2}$$

where

μ_α, μ_β are the normal deviates (z scores) corresponding to α and β;

D is the standard deviation.

For example, if α and β are both set to 0,05 and the aim is to detect a difference between the means of half a standard deviation:

$$N = \frac{2(1{,}65+1{,}65)^2}{0{,}5^2} = 87{,}12 \text{ rounded to } 87.$$

Hence, at least 87 test participants should be tested.

8.8.3 Statistical test

Barnard's U test [17] is used to compare the average search speeds and the ratings of visual comfort for the test and the benchmark displays. This test presupposes an *interval* scale for differences between test and benchmark values of search speed and visual comfort ratings. This is clearly true for search speed differences, whereas it is less trivial for visual comfort rating differences. The visual comfort ratings are done using a numerical category scale, i.e. an *ordinal* scale. However, it has been shown on several occasions that numerical category ratings are nearly linearly related to the corresponding values on the interval scale that is constructed with Thurstone's law of categorical judgment [18], [19], [20]. Hence, Barnard's U test can indeed be used for comparing the average search speeds as well as the ratings of visual comfort. Tables 2 to 4 provide a step-by-step guide to Barnard's U test, and a worked example.

Table 2 —Barnard's U test

Step	Description	Symbol/Equation
1	a) Record α, the risk of asserting a significant difference when the displays are the same, and β, the risk of asserting no significant difference when the displays are in fact different; both shall be set to 0,05.	α, β
	b) Record D, the difference — in units of standard deviation — between the means that is important to detect; it shall be set to 0,5.	D
2	For each test participant, obtain a score for the benchmark display (x_0) and for the test display (x_1).	x_0, x_1
3	Compute the score difference.	$x_0 - x_1$
4	Compute F, the sum of the score differences for all test participants tested.	$F = \sum (x_0 - x_1)$
5	Compute S, the sum of the squared differences.	$S = \sum (x_0 - x_1)^2$
6	Compute the U statistic.	$U = \frac{F}{\sqrt{S}}$
7	Compare this statistic with the boundary values, U_0 and U_1 according to the appropriate values of α, β and D (see 8.8.1). If $U < U_0$ then the null hypothesis is not rejected, and the test display passes. If $U > U_1$ then the null hypothesis is rejected in favour of the alternative hypothesis, and the test display fails. If $U_0 \leqslant U \leqslant U_1$, no decision can be made and testing shall continue.	—

Consider the worked example of Table 3, where x_1 and x_0 denote the average search velocities (in characters per second) for a test display and a benchmark display, respectively.

Table 3 — Example of sequential testing using Barnard's U test

N	x_1	x_0	$x_0 - x_1$	F	S	U	U_0 [a]	U_1 [a]
1	9,78	7,92	–1,86	–1,86	3,50	–1,000	—	—
2	17,19	14,48	–2,71	–4,57	10,8	–1,391	—	—
3	38,32	39,39	1,07	–3,50	12,0	–1,007	—	—
4	16,08	14,20	–1,88	–5,38	15,5	–1,364	—	—
5	13,56	12,17	–1,39	–6,76	17,4	–1,621	—	—
6	19,57	11,45	–8,12	–14,88	83,4	–1,629	–2,070	—
7	6,26	6,38	0,12	–14,76	83,4	–1,616	–1,790	—
8	8,20	7,06	–1,14	–15,90	84,7	–1,728	–1,510	2,560
9	24,16	22,23	–1,93	–17,83	88,4	–1,896	–1,330	2,510
10	10,35	7,90	–2,45	–20,28	94,4	–2,087	–1,150	2,460
11	13,83	10,37	–3,46	–23,74	106	–2,306	–1,034	2,436
12	12,21	6,97	–5,24	–28,98	134	–2,503	–0,918	2,412

N is the number of test participants.

For the meaning of the other symbols, see Table 2.

[a] The critical values are listed in Table 4.

After eight test participants, $U < U_0$; therefore the null hypothesis is not rejected, i.e. the search speed for the test display is not significantly slower than for the benchmark display and the test display passes this part of the test.

8.9 Critical values for Barnard's U test

Table 4 provides critical values for Barnard's U test for $\alpha = 0{,}05$, $\beta = 0{,}05$ and $D = 0{,}5$. These values are interpolated (using linear regression) from Table L.3 in Reference [16]. Boundary values, shown in Table 4 in square brackets, are included to assist in the drawing of boundaries and shall not be used in making a decision.

Table 4 — Critical values for Barnard's U test

Test participant	U_0	U_1
2	[–6,96]	—
3	[–5,045]	—
4	[–3,13]	[3,01]
5	[–2,6]	[2,87]
6	–2,07	[2,73]
7	–1,79	[2,645]
8	–1,51	2,56
9	–1,33	2,51
10	–1,15	2,46
11	–1,034	2,436
12	–0,918	2,412
13	–0,802	2,388
14	–0,686	2,364
15	–0,57	2,34
16	–0,498	2,334
17	–0,426	2,328
18	–0,354	2,322
19	–0,282	2,316
20	–0,21	2,31
21	–0,154	2,308
22	–0,098	2,306
23	–0,042	2,304
24	0,014	2,302
25	0,07	2,3
26	0,114	2,304
27	0,158	2,308
28	0,202	2,312
29	0,246	2,316
30	0,29	2,32
31	0,328	2,328
32	0,366	2,336
33	0,404	2,344
34	0,442	2,352
35	0,48	2,36
36	0,514	2,368
37	0,548	2,376
38	0,582	2,384
39	0,616	2,392
40	0,65	2,4
41	0,678	2,408
42	0,706	2,416
43	0,734	2,424
44	0,762	2,432

Test participant	U_0	U_1
45	0,79	2,44
46	0,816	2,45
47	0,842	2,46
48	0,868	2,47
49	0,894	2,48
50	0,92	2,49
51	0,944	2,499
52	0,968	2,508
53	0,992	2,517
54	1,016	2,526
55	1,04	2,535
56	1,064	2,544
57	1,088	2,553
58	1,112	2,562
59	1,136	2,571
60	1,16	2,58
61	1,18	2,59
62	1,2	2,60
63	1,22	2,61
64	1,24	2,62
65	1,26	2,63
66	1,28	2,64
67	1,3	2,65
68	1,32	2,66
69	1,34	2,67
70	1,36	2,68
71	1,378	2,69
72	1,396	2,70
73	1,414	2,71
74	1,432	2,72
75	1,45	2,73
76	1,468	2,74
77	1,486	2,75
78	1,504	2,76
79	1,522	2,77
80	1,54	2,78
81	1,557	2,79
82	1,574	2,80
83	1,591	2,81
84	1,608	2,82
85	1,625	2,83
86	1,642	2,84
87	1,659	2,85

Annex A
(informative)

Overview of the ISO 9241 series

This annex presents an overview of ISO 9241: its structure, subject areas and the current status of both published and projected parts, at the time of publication of this part of ISO 9241. For the latest information on the series, see:

http://isotc.iso.org/livelink/livelink?func=ll&objId=651393&objAction=browse&sort=name.

Part no.	Subject/title	Current status
1	General introduction	International Standard (intended to be replaced by ISO/TR 9241-1 and ISO 9241-130)
2	Guidance on task requirements	International Standard
3	Visual display requirements	Replaced by the ISO 9241 “300” subseries
4	Keyboard requirements	International Standard (intended to be replaced by the ISO 9241 “400” subseries)
5	Workstation layout and postural requirements	International Standard (intended to be replaced by ISO 9241-500)
6	Guidance on the work environment	International Standard (intended to be replaced by ISO 9241-600)
7	Requirements for display with reflections	Replaced by the ISO 9241 “300” subseries
8	Requirements for displayed colours	Replaced by the ISO 9241 “300” subseries
9	Requirements for non-keyboard input devices	International Standard (intended to be replaced by the SO 9241 “400” subseries)
11	Guidance on usability	International Standard
12	Presentation of information	International Standard (intended to be replaced by ISO 9241-111 and ISO 9241-141)
13	User guidance	International Standard (intended to be replaced by ISO 9241-124)
14	Menu dialogues	International Standard (intended to be replaced by ISO 9241-131)
15	Command dialogues	International Standard (intended to be replaced by ISO 9241-132)
16	Direct manipulation dialogues	International Standard (intended to be replaced by ISO 9241-133)
17	Form filling dialogues	International Standard (intended to be replaced by ISO 9241-134)
20	Accessibility guidelines for information/communication technology (ICT) equipment and services	International Standard

Part no.	Subject/title	Current status
Introduction		
100	Introduction to software ergonomics	Planned
General principles and framework		
110	Dialogue principles	International Standard
111	Presentation principles	Planned to partially revise and replace ISO 9241-12
112	Multimedia principles	Planned to revise and replace ISO 14915-1
113	GUI and controls principles	Planned
Presentation and support to users		
121	Presentation of information	Planned
122	Media selection and combination	Planned to revise and replace ISO 14915-3
123	Navigation	Planned to partially revise and replace ISO 14915-2
124	User guidance	Planned to revise and replace ISO 9241-13
129	Individualization	Planned
Dialogue techniques		
130	Selection and combination of dialogue techniques	Planned to incorporate and replace ISO 9241-1:1997/Amd 1:2001
131	Menu dialogues	Planned to replace ISO 9241-14
132	Command dialogues	Planned to replace ISO 9241-15
133	Direct-manipulation dialogues	Planned to replace ISO 9241-16
134	Form-based dialogues	Planned to replace ISO 9241-17
135	Natural language dialogues	Planned
Interface control components		
141	Controlling groups of information (including windows)	Planned to partially replace 9241-12
142	Lists	Planned
143	Media controls	Planned to partially revise and replace ISO 14915-2
Domain-specific guidance		
151	Guidance on World Wide Web user interfaces	International Standard
152	Interpersonal communication	Planned
153	Virtual reality	Planned
Accessibility		
171	Guidance on software accessibility	International Standard

Part no.	Subject/title	Current status
Human-centred design		
200	Introduction to human-centred design standards	Planned
210	Human-centred design of interactive systems	Planned to revise and replace ISO 13407
Process reference models		
220	Human-centred lifecycle processes	Planned to revise and replace ISO/PAS 18152
Methods		
230	Human-centred design methods	Planned to revise and replace ISO/TR 16982
Ergonomic requirements and measurement techniques for electronic visual displays		
300	Introduction to electronic visual display requirements	International Standard
302	Terminology for electronic visual displays	International Standard
303	Requirements for electronic visual displays	International Standard
304	User performance test methods for electronic visual displays	International Standard
305	Optical laboratory test methods for electronic visual displays	International Standard
306	Field assessment methods for electronic visual displays	International Standard
307	Analysis and compliance test methods for electronic visual displays	International Standard
308	Surface-conduction electron-emitter displays (SED)	Technical Report
309	Organic light-emitting diode (OLED) displays	Technical Report
Physical input devices		
400	Principles and requirements for physical input devices	International Standard
410	Design criteria for physical input devices	International Standard
411	Laboratory test and evaluation methods for the design of physical input devices	Planned
420	Selection procedures for physical input devices	Under preparation
421	Workplace test and evaluation methods for the use of physical input devices	Planned

Part no.	Subject/title	Current status
Workstation		
500	Workstation layout and postural requirements	Planned to revise and replace ISO 9241-5
Work environment		
600	Guidance on the work environment	Planned to revise and replace ISO 9241-6
Application domains		
710	Introduction to ergonomic design of control centres	Planned
711	Principles for the design of control centres	Planned to revise and replace ISO 11064-1
712	Principles for the arrangement of control suites	Planned to revise and replace ISO 11064-2
713	Control room layout	Planned to revise and replace ISO 11064-3
714	Layout and dimensions of control centre workstations	Planned to revise and replace ISO 11064-4
715	Control centre displays and controls	Planned to revise and replace ISO 11064-5
716	Control room environmental requirements	Planned to revise and replace ISO 11064-6
717	Principles for the evaluation of control centres	Planned to revise and replace ISO 11064-7
Tactile and haptic interactions		
900	Introduction to tactile and haptic interactions	Planned
910	Framework for tactile and haptic interaction	Under preparation
920	Guidance on tactile and haptic interactions	Under preparation
930	Haptic and tactile interactions in multimodal environments	Planned
940	Evaluation of tactile and haptic interactions	Planned
971	Haptic and tactile interfaces to publicly available devices	Planned

Bibliography

[1] ISO 9241-11, *Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability*

[2] ISO 9241-300, *Ergonomics of human-system interaction — Part 300: Introduction to electronic visual display requirements*

[3] ISO 9241-305, *Ergonomics of human-system interaction — Part 305: Optical laboratory test methods for electronic visual displays*

[4] ISO 9241-307, *Ergonomics of human-system interaction — Part 307: Analysis and compliance test methods for electronic visual displays*

[5] ISO 13407, *Human-centred design processes for interactive systems*

[6] ISO 20282-1, *Ease of operation of everyday products — Part 1: Design requirements for context of use and user characteristics*

[7] ISO/TS 20282-2, *Ease of operation of everyday products — Part 2: Test method for walk-up-and-use products*

[8] ISO/PAS 20282-3, *Ease of operation of everyday products — Part 3: Test method for consumer products*

[9] ISO/PAS 20282-4, *Ease of operation of everyday products — Part 4: Test method for the installation of consumer products*

[10] ISO/IEC 17025*, General requirements for the competence of testing and calibration laboratories*

[11] ISO/IEC 25062, *Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Common Industry Format (CIF) for usability test reports*

[12] ROUFS, J.A.J. and BOSCHMAN, M.C., Text quality metrics for visual display units: I. Methodological aspects. *Displays*, 18, pp. 37-43 (1997)

[13] BOSCHMAN, M.C. and ROUFS, J.A.J., Text quality metrics for visual display units: II. An experimental survey. *Displays*, 18, pp. 45-64 (1997)

[14] American Psychological Association (1990), Ethical principles of psychologists. *American Psychologist*, 45, pp. 390-395

[15] British Psychological Society (1991), *Code of Conduct, Ethical Principles and Guidelines*. Leicester: British Psychological Society

[16] HAYS, W. L., *Statistics*. New York: Holt, Rinehart and Winston (1963)

[17] BARNARD, G. A., Sequential tests in industrial statistics. *Journal of the Royal Statistical Society* B 8, pp. 1-21 (1946)

[18] THURSTONE, L.L., A law of comparative judgment*. Psychol. Review*, 34, pp. 273-286 (1927)

[19] ROUFS, J.A.J. and BOSCHMAN, M.C., Visual Comfort and Performance. Vision and Visual Dysfunction, Volume 15: *The Man-Machine Interface*, pp. 24-40. Edited by J.A.J. Roufs. Macmillan Press, London (1991)

[20] BOSCHMAN, M.C., ThurCatD: A tool for analyzing ratings on an ordinal category scale. *Behavior Research Methods, Instruments, & Computers*, 32 (3), pp. 379-388 (2000)

[21] DAVIES, O. L., *The Design and Analysis of Industrial Experiments*. London: Oliver and Boyd (1954)

ICS 13.180; 35.180
Price based on 21 pages