INTERNATIONAL STANDARD

ISO 9241-3

First edition
1992-07-15

AMENDMENT 1
2000-12-15

Corrected version
2002-12-15

Ergonomic requirements for office work with visual display terminals (VDTs) —

Part 3: Visual display requirements

AMENDMENT 1

Exigences ergonomiques pour travail de bureau avec terminaux à écrans de visualisation (TEV) —

Partie 3: Exigences relatives aux écrans de visualisation

AMENDEMENT 1

Reference number
ISO 9241-3:1992/Amd.1:2000(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

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 3.

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 Amendment may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

Amendment 1 to International Standard ISO 9241-3:1992 was prepared by Technical Committee ISO/TC 159, *Ergonomics*, Subcommittee SC 4, *Ergonomics of human-system interaction*.

This corrected version of ISO 9241-3:1992/Amd.1:2000 incorporates the following corrections.

On page 1, the phrase “Delete NOTE 11, in 7.1” has been deleted, indicating that the entire text of 7.1 has been replaced.

The text of 7.2 e) has now been deleted.

On page 13, a new sentence has been added before “Add the following references to the Bibliography”, and the publication date of [28] has been changed.

Ergonomic requirements for office work with visual display terminals (VDTs) —

Part 3:
Visual display requirements

AMENDMENT 1

Pages iii and iv, Contents

Replace the titles of annex C and clauses C.2 and C.5 to C.12 with the following:

Annex C Visual performance and comfort test

C.2 Test participants

C.5 Methods

C.6 Dependent measures

C.7 Statistical treatment of results

C.8 Conformance

C.9 Critical values for Barnard's U test

Page vi, Foreword

Replace the final sentence with the following text:

Annex C forms a normative part of this part of ISO 9241. Annexes A, B and D are for information only.

Pages 15 and 16, subclause 7.1

Replace the existing text of 7.1 with the following text:

7.1 Conformance with this part of ISO 9241 is achieved by meeting all the mandatory requirements of clause 5.

Where the physical measurements required to comply with clause 5 cannot be carried out because the display under test uses a new technology that makes the metrological measurements difficult or impossible, compliance is achieved by obtaining a positive result using the test method and associated mandatory requirements specified in annex C.

Page 16, subclause 7.2

Delete 7.2 e).

Page 22, annex C

Replace annex C with the following text.

Annex C
(normative)

Visual performance and comfort test

C.1 Principle

C.1.1 Purpose and use of test

This annex describes a procedure for testing the visual quality of VDTs where the entire set of physical requirements (defined in clause 5) cannot be applied, for example, to novel display technologies such as new types of flat panel display (such as electroluminescent and field emission displays). The test consists of a search task and an assessment of visual comfort. The combination of test results will be referred to as the visual quality of a display.

C.1.2 Intent and scenario of use

The intention behind this test method is to provide a test method for displays that cannot otherwise be tested for conformance with this part of ISO 9241. This test method is not an alternative test method, in the sense that a display manufacturer can choose either the physical requirements specified in clause 5 or this method. Instead, the test method provides a testing route for displays that *cannot be tested* according to the requirements specified in clause 5 because the display under test uses a new technology that makes the metrological measurements difficult or impossible (see 7.1).

C.1.3 Overview of test method

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 and business graphics. The dependent variables of the test are the search velocity achieved by the test participants in a visual letter search task and subjective ratings of visual comfort using a category scale. The combination of the test results will be referred to as the visual quality of a display. 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 IPO, Center for Research on User-System Interaction (Boschman & Roufs [3]).

The visual quality of a display, referred to as a test display, is assessed against a benchmark display known to meet or exceed the mandatory requirements of clause 5 of this part of ISO 9241. Both the velocity 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 or falls short of performance on the benchmark display.

C.1.4 Avoidance of bias

All tests are open to bias, and this is especially true in the area of psychological testing. The assessment should therefore be carried out under the supervision of those qualified to carry out such testing, with the necessary education and at least one year of experience. Rules governing the ethical conduct of human experimental testing

should be followed. Examples of such rules can be found in the American Psychological Association (1990) [23] and the British Psychological Society (1991) [27].

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; the list is not intended to be complete.

— Selection of test participants (e.g. avoid selection of particular age groups).

— Configuration of displays (during the test, the benchmark display shall meet all the requirements of clause 5).

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

— Instructions to the test participants (these should be impartial).

C.2 Test participants

Test participants should be a sample representing the anticipated user population (those who perform office tasks as specified in the scope of this part of ISO 9241). All test participants shall have near visual acuity that is normal, or corrected to normal, at the design viewing distance and shall be without any obvious physical or physiological conditions that could influence either their search performance or their perceived image quality.

C.3 The displays

The test display shall be a production or full-feature pre-production unit. It shall incorporate all anti-glare and reflection filters and treatments that will be in the production unit. The benchmark display shall be supplied or nominated by the supplier of the test display and shall meet or exceed all mandatory requirements in clause 5 of this part of ISO 9241.

The displays may be labelled for identification purposes (e.g. "Display 1" and "Display 2"). Under these conditions, the test participants should not be informed which is the test and which is the benchmark display, so half of them should have the test display labelled as “Display 1” and the other half should have the benchmark display labelled as “Display 1”.

C.4 Test workstation and environment

C.4.1 General requirements

The test shall be conducted in an area that is free from distractions and external interference which 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, both within a test participant’s session and between test participants.

C.4.2 Environment

The thermal environment, the background noise level, the ambient lighting, and the reflectance of work surfaces shall meet the minimum requirements in ISO 9241-5 and ISO 9241-6. The ambient illumination shall be designed to minimize glare and specular reflections (see ISO 9241-7). 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 may be used by the experimenter to instruct each test participant about the test.

C.4.3 Workstation for the test

The display and associated equipment (for example, the keyboard) shall be supported by a work surface that meets the requirements of ISO 9241-5.

For both the benchmark display and the test display, the viewing distance shall be set according to their 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 in 5.2 of this part of ISO 9241.

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 mandatory requirements in clause 5 of this part of ISO 9241.

NOTE A measurement procedure for specifying brightness and contrast is given in this part of ISO 9241.

According to the manufacturer's wishes, the brightness and contrast settings 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 the test.

The test participant shall be seated in a chair that meets the requirements of ISO 9241-5.

C.5 Methods

C.5.1 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 given in 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" ... "Z", "a" ... "z", and "0" ... "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-texts shall consist of blocks of random strings of characters separated by spaces.

— The texts, on both test and benchmark display, 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 (see C.5.2) shall be sized such that, if 5 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 is 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 may not start or end with the target character.

— The texts 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). Although the average word length does vary over different languages, pseudo-texts with 15 % space fraction, in a way, 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 2 characters.

C.5.2 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 the presence of the target character.

Place the blocks of pseudo-text in the upper left, the upper right, the lower left, the lower right and the 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.

Tell 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 controls 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 may give the user an indication of the display under test and therefore may affect the results of the test. This can be prevented by asking the user to adjust the controls 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 the 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 the presence of the target character in each block of pseudo-text and the number of errors made by the test participant (see clause C.6). 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 work as quickly as possible and to minimize errors as far as possible.

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, 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. The scale to be used is specified below.

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. You should circle the number corresponding to your judgement."

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

NOTE An example of a set of instructions for test participants is given in C.5.4.

C.5.3 Task conditions

Display attributes (character size, resolution, visual angle, fonts and so on) 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 mandatory requirements defined for size, shape and spacing of characters in 5.4, 5.5, 5.6, 5.8, 5.9, 5.10, and 5.11.

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

— A target character shall have intermediate discriminability with the other characters used (e.g. 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 about 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 which 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.

C.5.4 Instructions to test participants

These are sample instructions that should be modified for your own testing situation, e.g. these instructions 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. They will need to be modified if a non-keyboard input device or another direction of reading is used.

The instructions shall be presented to the participant on paper, an example of how they may 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 quickly and as accurately 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 quickly and accurately as possible. 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 Oulp2JBE
vbR l1Y5rVr SA9mr DmPETLV 2uO2
7phnFd2oyT 83ee zKo8n 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
fySFoaWxj 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
```

C.6 Dependent measures

C.6.1 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 velocity obtained from trials with error rates < 10 %;

b) the subjective ratings of visual comfort.

Error rate, E, is defined as:

$$E = \frac{|T_o - T_c|}{T_o} \times 100\ \%$$

where

T_o 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 (1 missed or extra target is accepted in a text with 10 targets.

C.6.2 Average search velocity

From the registered search times T_i corresponding with the valid trials (E < 10 %) the performance measure of a test participant, the *average search velocity*, v_s, measured in character(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 may be analysed by applying a sequential testing procedure for successive test participants (see clause C.7).

C.6.3 Subjective ratings

Each test participant shall give, for both the test and the benchmark display, a subjective rating of visual comfort on a 9-point scale.

NOTE These ratings may be analysed using the sequential testing procedure for successive test participants described in C.7.

C.7 Statistical treatment of results

C.7.1 General

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

NOTE 1 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.

NOTE 2 Other statistical procedures and analysis may be carried out as long as they are of comparable robustness.

If the sequential analysis procedure below is not used, the test and statistical analysis shall ensure that the Type 2 error rate is $\leqslant$ 0,05 for a standard deviation, D, of 0,5, and ensure that the criterion α (manufacturer's risk) shall be 0,05 (see Table C.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 if performance for the test product is significantly worse 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.

C.7.2 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 of error (Type 2) occurs when the null hypothesis is falsely not rejected. These two risks are usually symbolized by α and β (see Table C.1).

Table C.1 — The 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 Hays [31].

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

where

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

D is the standard deviation.

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

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

and hence at least 87 test participants should be tested.

C.7.3 Statistical test

Barnard's U test (Barnard, 1946) is used to compare the average search velocities and the ratings of visual comfort for the test and the benchmark display. Tables C.2 to C.4 provide a step-by-step guide to Barnard's U test, and a worked example.

Table C.2 — Description of Barnard's U test

Step	Barnard's U Test	
1	(i) 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, shall both be set to 0,05 (ii) Record D, the difference – in units of standard deviation – between the means that is important to detect shall be set to 0,5.	α, β 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 difference score.	$x_0 - x_1$
4	Compute F, the sum of the difference scores 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	This statistic is then compared with boundary values, U_0 and U_1 according to the appropriate values of α, β and D (see clause C.10). 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 must continue.	

For example, consider the following worked example, where x_1 and x_0 denote the average search velocities (in characters per second) for a test display and a benchmark display, respectively.