TECHNICAL SPECIFICATION

ISO/TS 20282-2

Second edition
2013-08-01

Usability of consumer products and products for public use —

Part 2: Summative test method

Facilité d'emploi des produits de consommation courante et des produits à usage public —

Partie 2: Méthode d'essai sommative

Reference number
ISO/TS 20282-2:2013(E)

COPYRIGHT PROTECTED DOCUMENT

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.

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. 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. 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 159, *Ergonomics*, Subcommittee SC 4, *Ergonomics of human-system interaction*.

This second edition cancels and replaces the first editions of ISO/TS 20282-2:2006, ISO/PAS 20282-3:2007, and ISO/PAS 20282-4:2007, of which it constitutes a technical revision. The main changes are the following:

— clarification that the test method is only intended to be used when there are a limited number of goals to be tested and it is possible to identify typical contexts of use and criteria for successful goal achievement;

— provision for a wider range of methods to test achievement of goals;

— use of ISO 9241-11 terminology: usability, effectiveness, efficiency, and satisfaction;

— wider range of levels of confidence with a sample size as low as 10, using the Adjusted Wald statistic;

— wider range of purposes for use of the test method.

ISO/TS 20282 consists of the following parts, under the general title *Usability of consumer products and products for public use*:

— *Part 1: Design requirements for context of use and user characteristics*

— *Part 2: Summative test method* (Technical Specification)

Introduction

Many people find some consumer products and walk-up-and-use products, including consumer products provided for public use, difficult to install and use, particularly when using them for the first time or at infrequent intervals. This is clearly undesirable for the producers of such products, for organizations that use the products to provide a service, and for the people who use them. Information about the usability of a product would, therefore, be of great value to producers, as part of development and marketing, to service providers, and to potential purchasers making purchase decisions or comparing alternative products. This would provide an incentive for producing products that are easier to install and use and would enable potential purchasers to pay specific attention to usability when selecting a product to buy and use. It is difficult to judge usability in a purchase situation without available comparable usability test results.

Usability (see ISO 9241-11) is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use. Effectiveness is fundamental as it is about achieving the intended goal(s). Efficiency is about the resources (such as time or effort) needed by users to achieve their goals so it can be important. In addition, it is important that users are satisfied with their experience, particularly where users have discretion over whether to use a product and can readily choose some alternative means of achieving their goals. In this part of ISO/TS 20282, accessibility is operationalized as the extent to which a product can be used with effectiveness, efficiency, and satisfaction by people from a population with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use.

Poor usability and/or accessibility can result in errors that can lead to several types of risks, for example, inconvenience resulting from not achieving a goal or achieving the wrong goal, incurring unexpected costs, or physical injury. In many countries, there are legal requirements to provide accessible products, services, and facilities.

EXAMPLE Calling the wrong person by mistake with a mobile phone may have the negative consequence of possible undesirable call charges either for the caller or the person called (who may have to pay for the call).

In addition to the risks of potential adverse consequences for the user as a result of failing to achieve their goal or achieving the wrong goal (poor effectiveness), there are other risks such as being late as a result of poor efficiency or users avoiding the use of a difficult-to-use product as a result of poor satisfaction.

Formative evaluation using expert inspection or user-based testing to provide feedback to improve the usability of the product is an integral part of the iterative human-centred design process recommended in ISO 9241-210. Summative evaluation can be used to validate usability and/or accessibility requirements, to provide a benchmark, or to provide a basis for comparison of different products. Although some types of expert inspection methods based on a checklist or a standard can provide summative data, the aspects of usability and/or accessibility that are measured are limited in comparison with the measures of effectiveness, efficiency, and satisfaction provided by user-based testing.

EXAMPLE One study found that only 50 % of the problems encountered on 16 websites by 32 blind users were covered by Success Criteria in the Web Content Accessibility Guidelines 2.0 (WCAG 2.0).[23]

Inspection can precede user-based testing to identify (and, if possible, eliminate) easily identifiable problems and to check that the product is capable of achieving the intended goals for the intended users (see 7.4).

To provide reliable data on effectiveness, efficiency, and satisfaction that can be compared, it is desirable to have a standard summative user-based test procedure. This part of ISO/TS 20282 specifies a summative user-based test method that can be used to provide an evaluation of the usability and/or accessibility and ease of unpacking, setting up, and installation of consumer products, and the usability and/or accessibility of products for public use (including walk-up-and-use products). It can be applied to products that are used to achieve goals that have clear success criteria and relate to well-defined types of subject matter.

ISO/TS 20282-1 describes in more detail sources of variance in user characteristics that form part of the context of use that needs to be taken into account when designing for usability. This information is also

needed to identify the elements of the context of use required for testing in this part of ISO/TS 20282. Further information about the characteristics of older people and people with disabilities can be found in ISO/TR 22411.

Usability of consumer products and products for public use —

Part 2: Summative test method

1 Scope

This part of ISO/TS 20282 specifies a user-based summative test method for the measurement of the usability and/or accessibility of consumer products and products for public use (including walk-up-and-use products) for one or more specific user groups. This test method treats accessibility as a special case of usability where the users taking part in the test represent the extremes of the range of characteristics and capabilities within the general user population. When the test method refers to usability, the method can also be used to test accessibility (unless otherwise specified).

This test method is for use when valid and reliable measures of effectiveness, efficiency, and satisfaction are needed.

NOTE 1 Products for public use include walk-up-and-use products that provide a service to the general public.

The test method can also be used to assess the usability and/or accessibility of achieving the goals of unpacking, installing, and setting up a consumer product.

This part of ISO/TS 20282 is intended to be used for testing the usability and/or accessibility of products when

— it is possible to identify typical contexts of use that are representative of the use of the product(s),

— it is possible to identify the criteria for the successful achievement of the users' goal, and

— there are a limited number of goals being tested at the same time.

While the test method is intended to test consumer products and products for public use, it can also be used to test other products, systems, and services with the characteristics described above.

If use of a product involves interaction with inputs, outputs, or environments that are highly variable and/or complex with variability or complexity that cannot be categorized in well-defined subsets, it is outside the scope as it would not be possible to obtain reliable results. See Annex A for examples of products and goals that are within the scope of this part of ISO/TS 20282.

EXAMPLE The method could be applied to an office photocopier, a website selling books or train tickets, or a legal advice service. The method would not be appropriate for a complex ecommerce website, a word processor, or a bicycle.

The method is primarily intended for use for assessing completed versions of products, but could also be used for internal purposes during development to judge, assess, and communicate the usability and/or accessibility of functional prototype versions.

The results of the summative test method can be used for the following purposes:

— to estimate the probability of achieving target values of effectiveness, efficiency, and satisfaction in actual use;

— to publish information about the usability and/or accessibility of a product;

— to compare the usability and/or accessibility of several products;

— to compare the results with a usability and/or accessibility requirements specification;

— to support procurement.

NOTE 2 Annex H lists the information to be included when specifying the procedure used to test whether the usability and/or accessibility requirements (Annex G) have been met.

The intended users of this part of ISO/TS 20282 are people with expertise in the design and management of testing usability and/or accessibility, working within or on behalf of manufacturers, suppliers, purchasing organizations, or third parties (such as test organizations or consumer organizations).

2 Conformity

A report of the values for the usability and/or accessibility of a product conforms to this part of ISO/TS 20282 if

— the test method used conforms to the requirements in Clauses 6, 7, 8, and 9 and Annexes C and D, and

— the report of the results contains the information specified in Annex F.

A statement of requirements for usability results conforms to this part of ISO/TS 20282 if it conforms to the requirements in Annex G.

The specification of a usability test procedure conforms to this part of ISO/TS 20282 if it conforms to the requirements in Annex H.

3 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

4 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

4.1
accessibility
extent to which products, systems, services, environments, and facilities can be used by people from a population with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use

Note 1 to entry: Context of use includes direct use or use supported by assistive technologies.

[SOURCE: ISO 26800:2011, definition 2.1]

Note 2 to entry: When carrying out a summative user-based measurement of accessibility, all three components of usability (effectiveness, efficiency, and satisfaction) need to be considered.

4.2
actual users
group(s) of people who directly interact with a product

Note 1 to entry: Before a product is released, this is the intended user group, and after release this is based on what is known about the actual user group.

[SOURCE: ISO 20282-1:2006, definition 3.1]

4.3
consumer product
product that is intended to be acquired and used by an individual for personal rather than professional use

[SOURCE: ISO 20282-1:2006, definition 3.2]

4.4
context of evaluation
users, tasks, equipment (hardware, software, and materials), and the physical and social environments in which a product is evaluated

[SOURCE: ISO/TS 20282-2:2006, definition 4.3]

4.5
context of use
users, tasks, equipment (hardware, software, and materials), and the physical and social environments in which a product is used

[SOURCE: ISO 9241-11:1998, definition 3.5]

4.6
ease of interaction
usability of interaction with the user interface of a product

Note 1 to entry: Ease of interaction is the effectiveness, efficiency, and satisfaction with which the user can successfully interact with the interface of the product.

4.7
effectiveness
accuracy and completeness with which users achieve specified goals

[SOURCE: ISO 9241-11:1998, definition 3.2]

4.8
efficiency
resources expended in relation to the accuracy and completeness with which users achieve goals

[SOURCE: ISO 9241-11:1998, definition 3.3]

4.9
formative evaluation
evaluation designed and used to improve the object of evaluation, especially when it is still being developed

[SOURCE: ISO/TR 18152:2010, definition 4.6]

Note 1 to entry: A formative test method is used to perform a formative evaluation.

4.10
goal
intended outcome

[SOURCE: ISO 9241-11:1998, definition 3.8]

Note 1 to entry: A goal is stated independently of the functionality used to achieve it.

4.11
intended users
group(s) of people for whom a product is designed

Note 1 to entry: In many cases, the actual user population is different from that originally intended by the manufacturer. The intended user group is based on realistic estimations of who the actual users of the product will be.

[SOURCE: ISO 20282-1:2006, definition 4.12]

4.12
interaction
bidirectional information exchange between users and equipment

[SOURCE: IEC/TR 61997:2001, definition 3.4]

Note 1 to entry: Equipment includes both hardware and software.

Note 2 to entry: Information exchange may include physical actions, resulting in sensory feedback.

4.13
main goal(s)
most frequent or important goal(s) that all or a large majority of users want to achieve when using a product

[SOURCE: ISO 20282-1:2006, definition 3.14]

Note 1 to entry: Main goals can depend on achieving sub-goals.

Note 2 to entry: Examples of main goals are given in Annex A.

4.14
satisfaction
freedom from discomfort, and positive attitudes towards the use of the product

[SOURCE: ISO 9241-11:1998, definition 3.4]

4.15
stratified sample
sample established by a procedure in which the population is divided into subpopulations (strata), each one of which contributes with a specified number of randomly selected individuals

[SOURCE: ISO 15535:2006, definition 3.4]

4.16
success rate
<usability> percentage of users successfully achieving a goal

4.17
summative evaluation
evaluation designed to present conclusions about the merit or worth of the object of evaluation

Note 1 to entry: The results can be used to produce recommendations about whether it should be retained, altered, or eliminated.

Note 2 to entry: It is possible to design a method to provide a combined formative and summative evaluation.

Note 3 to entry: A summative test method is used to perform a summative evaluation.

[SOURCE: ISO/TS 18152:2010, definition 4.10, modified — Reference to recommendations has been removed and notes to entries have been added.]

4.18
task
activities required to achieve a goal

Note 1 to entry: These activities can be physical and/or cognitive.

[SOURCE: ISO 9241-11:1998, definition 3.9]

4.19
usability
extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use

[SOURCE: ISO 9241-210:2010, definition 2.13]

Note 1 to entry: References to the term "usability" in the description of the user-based test method include the application of the test method to accessibility (unless otherwise specified).

4.20
usability requirement
required level of usability expressed in terms of measures of effectiveness, efficiency, and satisfaction in a specified context of use

4.21
usability testing
evaluation that involves representative users performing specific tasks with the system to enable the measurement of efficiency, effectiveness, and/or user satisfaction

[SOURCE: ISO/IEC 25060:2010, definition 2.17]

4.22
user
person who interacts with a system, product, or service

Note 1 to entry: The person who uses a service provided by a work system, such as a customer in a shop or passenger on a train, can be considered a user.

[SOURCE: ISO 26800:2011, 2.10, modified — Notes 1 and 3 have been omitted.]

4.23
user characteristics
attributes of a user that could influence usability

[SOURCE: ISO 20282-1:2006, 3.20, modified — The word "may" has been replaced with "could".]

4.24
user group
<usability> group of users differentiated by characteristics of the users, tasks, or environments that are expected to influence usability

Note 1 to entry: This could either be an intended user group or a user test group.

4.25
user interface
elements of a product used to control it and receive information about its status

Note 1 to entry: A list of operating instructions permanently displayed on the product is part of the user interface.

EXAMPLE The user interface of a shower tap is the water control lever, where the movement of the lever controls the temperature of the water and the position of the lever communicates the temperature to the user.

[SOURCE: ISO 20282-1:2006, definition 3.21, modified — The phrase "and the interaction that enables the user to use it for its intended purpose" has been deleted.]

4.26
user test group
group of persons selected to participate in a test of usability, sampled according to specific requirements

4.27
walk-up-and-use product
product that provides a service to the general public

Note 1 to entry: Walk-up-and-use products are designed to enable users to approach and use the product successfully without previous experience.

5 General principles

5.1 Type of usage of the product to be tested

This user-based test method can be used to measure the usability and accessibility of

— unpacking, installation, and setting up of consumer products,

— use of consumer products,

— products for public use, including walk-up-and-use products that provide a service to the general public, and

— other products that are used to achieve goals that have clear success criteria and relate to well-defined types of subject matter.

NOTE Accessibility is particularly important for walk-up-and-use products and for consumer products provided for public use.

The test method measures effectiveness, efficiency, and satisfaction when used by specified user groups in specified contexts of use. Accessibility is measured by the extent to which products, systems, services, environments, and facilities can be used by people from a population with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use.

5.2 Purpose of the test

This part of ISO/TS 20282 can be used by product manufacturers to

a) test a single product to determine whether usability and/or accessibility requirements have been met,

b) test a single product to provide evidence of the usability and/or accessibility of a product for a customer or for marketing purposes,

c) test a single product to establish a benchmark against which future products can be compared,

d) make comparisons between different products,

e) make comparisons between versions of the same product, and

f) specify usability and/or accessibility requirements for a product to be developed (Annex G) and the scenarios to use when testing whether the requirements have been met (Annex H).

This part of ISO/TS 20282 can be used for procurement by corporate or public purchasers to

a) test a single product to decide whether it meets usability and/or accessibility requirements,

b) test a group of similar products to make comparisons to facilitate decisions on which is the most suitable,

c) ensure that tests of similar products use the same methodology so that information about usability and/or accessibility can be compared during the procurement process, and

d) specify usability and/or accessibility requirements for a product to be procured (for example, when products are to be used in a school, hotel, or homes for the elderly) (Annex G) and the scenarios to be used when testing whether the requirements have been met (Annex H).

NOTE 1 ISO/IEC 25062 provides a format for reporting this information for professional products.

This part of ISO/TS 20282 can be used by a third-party test organization or consumer advocate organization to

a) test a single or multiple product(s) to decide whether they meet the usability and/or accessibility requirements for a user group,

b) test a single or multiple product(s) to establish a benchmark against which future products can be compared,

c) compare alternative products to provide information to be included in reports, and

d) make comparisons between competing systems (testing the main goals of use of the type of product or system unless there are specific reasons for testing other goals).

EXAMPLE 1 A testing organization wants to find out how easy it is for new owners of mobile phones to make a call from the address book. The requirement is to be 80 % confident of an 80 % success rate. A representative group of 12 people who have not previously owned this brand of mobile phone are selected. To meet the requirement, 11 out of 12 users will have to be successful within a predefined time limit. Success rate and task time are recorded, and satisfaction is measured after each task using the system usability scale (SUS).

EXAMPLE 2 A manufacturer wants to demonstrate that a personal video recorder (PVR) is easy to program. The manufacturer has defined a requirement to be 95 % confident that 80 % of the user population can program the PVR. A representative sample of 30 people who have purchased a PVR with the intention of programming it to record programmes is recruited. They are selected to be representative in age, education, and ownership of different brands of PVR. Each session is expected to take a total of 15 min to 20 min and two PVRs are set up with TVs for programming. Each person is asked to program the PVR to record a specified programme. The test result is given as the measured success rate together with a confidence interval. The requirement is met when 28 of the 30 people tested are successful.

EXAMPLE 3 A manufacturer wants to demonstrate that a multifunction home printer/copier is easy to install for current users of other printers/copiers. The manufacturer has defined a requirement to be 95 % confident that 80 % of the user population can successfully install the printer. A sample of 30 people who own a printer/copier at home is recruited. They are selected to be representative in age, education, and ownership of different brands. Each session is expected to take a total of 10 min to 30 min and two kits are available. The printer/copier is returned to its default settings on completion of each test and repackaged as it would be received by the consumer. Each person is asked to install the printer/copier and print and copy a test page. The test result is given as the measured success rate together with a confidence interval and the geometric mean task time together with a confidence interval.

5.3 Scope of the goals used in the test

5.3.1 Selection of goals

Two decisions need to be made about the scope of the goals used in the test.

a) Which goal(s) are to be tested for achievement in which context(s) of use (see 5.3.2)?

b) Is goal achievement going to be assessed by whether interaction with the interface is successful (ease of interaction) or by whether the overall outcome resulting from the interaction meets the needs of the user (overall usability of the product)(see 5.3.3)?

5.3.2 Testing achievement of the main goals in the main context of use or specific goal(s) in specific context(s) of use

The scope of the test can be to

a) test the achievement of the main goals in the main context(s) of use in order to obtain measures of effectiveness, efficiency, and satisfaction that are representative of the overall usability or accessibility of the product. (When testing installation and setup, the main goal is to install the

most common configuration to obtain measures that are representative of the overall usability of the installation process.)

b) test the achievement of more specific goal(s) and/or using more specific context(s) of use to obtain measures that are representative of the usability of the product or installation process in specific context(s) of use (for example, for a specific user group (see 7.3.3), specific environment, or specific task).

5.3.3 Measuring installation and setup, the usability of the user interface, and interaction or overall usability

The prerequisites for usability of a product (i.e. the prerequisites for users to be effective, efficient, and satisfied) are

a) the specified users can successfully interact with the interface of the product in specified contexts of use (ease of interaction), and

b) the product is capable of producing results of acceptable technical quality as a result of the interaction.

While interaction with the product should enable the user to achieve their overall goals, the product also has to be capable of producing adequate results.

EXAMPLE 1 If the goal is to use a camera to take photographs that can be used for large prints, the user has to be able to successfully interact with the interface and the quality of the image generated by the camera needs to be adequate for a large print.

The scope of the test can be

a) unpacking, setting up, and installation of consumer products,

b) usability of the user interface and interaction ("ease of interaction"), or

NOTE If a product reliably produces results of acceptable technical quality, it is sufficient to just test the ease of interaction.

c) usability of the product as a whole ("overall usability": the extent to which the product enables the user to achieve their overall goal with effectiveness, efficiency, and satisfaction).

EXAMPLE 2 The effectiveness of achieving the goal of using an alarm clock to be woken in the morning could be decomposed.

Scope of test	Interactions	Goals (success criteria)	Assumptions
Unpacking, setting up, and installation	Unpack the clock and set the current time	Clock unpacked, working, and set to the correct time	None
Usability of the user interface and interaction (ease of interaction)	Set the alarm hours and minutes for a specified time and turn on the alarm on the clock	Alarm is correctly set (usability of the user interface)	When the alarm is correctly set, it will keep time and always sound at the specified time. Alarm is sufficiently loud to wake the user.
Usability of the product as a whole	Not specified	Alarm sounds at the specified time and wakes the user	None

The usability of an alarm clock for the goal of being woken in the morning could be tested by finding out whether, after setting the alarm, the user is woken at the specified time in the morning.

Alternatively, if it is known that when the alarm is correctly set, it will always sound at the specified time and will be sufficiently loud to wake the user, measuring the usability of interaction with the user

interface by observing whether the user correctly sets the time and turns the alarm on is sufficient to be confident of the usability of the product for the goal of being woken.

So if the alarm clock is known to have the required technical quality, testing the usability of interaction with the user interface is also a test of the usability of the product. If the technical quality of the alarm clock is not known, the usability of interaction with the user interface does not provide a guarantee of the usability of the alarm clock for waking the user up.

5.4 Validity and reliability

The design of user-based tests should pay particular attention to ensuring that the results will be

— valid, i.e. that the data that have been gathered are, in fact, measures of what the test intended to measure (i.e. they are appropriate measures of usability), and

— reliable, i.e. that the test would produce consistent results if it were repeated.

A valid and reliable test provides the basis from which one can justify that the results can be generalized to predict the usability of the tested product.

The test method specified in Clause 6 incorporates good experimental practice designed to maximize the validity and reliability of the results. It is particularly important that a representative sample of users takes part in the test if the results are to apply to the intended user population.

5.5 Ethical issues

Any test involving human participants should implement good practice in addressing ethical issues.

NOTE 1 Some countries have human subject research legislation. Many organizations follow established procedures and regulations. A number of professional organizations have codes of practice for the ethical governance of experimental work. The aim of such codes of practice is to safeguard the welfare and rights of those involved, including the participants and other people who could be affected by the test, as well as the investigators.

In particular, the design and conduct of tests should ensure that the rights and safety of individuals are protected. For example:

— No test should demand excessive levels of physiological or mental effort.

NOTE 2 This is of particular importance when tests involve participation by older people or people with disabilities.

— The privacy of the individuals involved should be safeguarded.

— The confidentiality of any data collected should be maintained.

— If some user groups are thought to be vulnerable, for example, children, older people, or people with disabilities, special care should be taken to establish that their rights have been protected.

— Participants should be provided with sufficient information about the purposes and conduct of any study to enable them to decide whether to participate or not on the basis of this information.

— Participants should be entitled to withdraw from participation at any stage.

6 Test procedure overview

To measure usability and/or accessibility using the summative test method in Clause 7, the following activities shall be carried out.

NOTE As there are interdependencies, some of the activities may need to take place in parallel.

a) Define the purpose and scope of the test (see 7.1).

b) Identify the product(s) to be tested (see 7.2).

c) Specify the goals to be included in the test (see 7.3.2).

d) Specify the user group(s) to be used for testing (see 7.3.3).

e) Consider whether to use the selection of user test groups to assess accessibility (see 7.3.3.4).

f) Specify environmental characteristics (see 7.3.4).

g) Check that product is compatible with the user characteristics and goals (see 7.4).

h) Identify the relevant measures (see 7.5.2).

i) Determine the number of participants required for the desired confidence level (see 7.5.3).

j) Recruit a sample of users (see 7.5.4).

k) Configure the product (see 7.5.5).

l) Specify test scenarios and goal criteria (see 7.5.6).

m) Establish the test procedure (see 7.5.7).

n) Create the test environment (see 7.5.8).

o) Assess whether the criteria for goal achievement have been met by individual users (see 7.5.10).

p) Calculate the results (see Clause 8).

q) Prepare a full report and, if required, a short summary (see Clause 9 and Annex F).

7 Test method

7.1 Purpose and scope of the test

The type of usage of the product to be tested shall be identified and, whether testing for accessibility is included, shall be decided (see 7.3.3.4). The purpose of the test (see 5.2) and scope (see 5.3) of the test shall be defined to enable an appropriate test to be designed that provides the required information.

7.2 Product(s) to be tested

Identify the specific product(s) and version(s) to be tested and, if appropriate, the organization responsible for the provision of the service supported by the product.

If any of the products to be tested do not have product certification, it will be necessary to ensure that they are safe for the participants to use.

EXAMPLE Issues could include electrical safety, sharp edges, or small loose parts when testing with children.

7.3 Context of evaluation

7.3.1 General

The context used for evaluation shall be specified. This context and the goals to be achieved should be as close as possible to the actual or intended goals and context of use for specific user group(s).

NOTE 1 More information about the context of use can be found in ISO/IEC 25063.

NOTE 2 If detailed information is not already available about the current or intended context of use, this will have to be obtained. For example, if a test is being carried out by or on behalf of a manufacturer, the manufacturer might already have this information.

7.3.2 Specify how goal achievement will be measured

7.3.2.1 Scope of goals

The measurement of usability can focus on (1) whether the users can successfully interact with the interface of the product, or (2) whether the product as a whole meets the usability criteria, or (3) whether the installation and setting up of a consumer product meets the usability criteria (see 5.3).

7.3.2.2 Specify the users' goals

The main purpose(s) of use of the product shall be identified.

The user goals to be tested shall be defined. The goals need to be defined in conjunction with the identification of user groups (see 7.3.3). Different user groups can have different goals.

For the goals to be used for testing, it shall be possible to

— identify typical contexts of use that are representative of the use of the product for that goal, and

— identify criteria for successful achievement of the goal.

The achievement of a goal can depend on the achievement of sub-goals.

EXAMPLE 1 The goal of watching news on TV could depend on achieving the sub-goals of turning the TV on and selecting an appropriate channel.

There is sometimes more than one way of achieving the same goal.

EXAMPLE 2 A recording could be located either by date and time or by name.

During first use of a product, installation can be a main goal. Once installation is complete, it would no longer be a main goal.

The users' goal(s) and the context of use of the product shall be identified based on an understanding of the users' needs, for example, based on information gained from the users, customer, or supplier. When testing the main goals (see 5.3.1), the goals that are tested shall include the most frequent and/or important types of users and user goals and the statement of results shall explain how the main goals were identified.

NOTE 1 When testing a walk-up-and-use product, the organization responsible for the operation of the product or the provision of the service might be able to provide a breakdown of the frequency and characteristics of the transactions expected to be performed using the product.

The goals and context of use to be tested shall be for outcomes that a user would reasonably expect to be able to achieve in that context of use when using the product for one or more of the purposes specified by the manufacturer.

If when testing several products one or more of the products does not provide functionality to support one or more goals or contexts of use to be used for testing, these goals or context(s) of use shall not be included in the test tasks used for these particular product(s), but information about the lack of support shall be included in the report.

NOTE 2 This situation can occur when carrying out comparative tests using several products with different ranges of functionality.

EXAMPLE One of the identified main goals of use of a mobile phone is being able to use a phone in a wide range of environments, including a moderately noisy environment, so this environment could be included in a comparative test even if the sound quality of one phone was insufficient for it to be audible in this environment. The phone that does not have sufficient sound quality is not to be included in the tests task for that context, but its inadequacy is reported.

7.3.2.3 Establish criteria for test goal achievement

The test goals shall be expressed in terms of the intended outcomes. When measuring overall usability, the goals shall be described independently of the means by which they are achieved.

EXAMPLE 1 A goal for an oven is to have the oven heated to a specific temperature. A goal for a ticket machine could be to have obtained the cheapest return ticket from London to Manchester valid for the desired dates and times of travel.

The criteria for adequately complete and accurate goal achievement shall be specified in a verifiable way. The criteria for what constitutes successful achievement of a goal shall as closely as possible match the criteria that would be applied by a typical user, rather than any constraints of the technology. For some products, there might be only one acceptable result (for example, to have a package opened). For others, there could be a range of outcomes that are acceptable (for example, accuracy of the temperature of an oven or dryness of clothes).

If an overall goal can be partially achieved, the criteria for achievement of the sub-goals that contribute to the overall goal shall be specified.

EXAMPLE 2 A goal when using a camera could be to have a photo of a distant object in low lighting. The best photo would be obtained using full zoom, anti-shake, and no flash. Not making one of these settings could reduce the quality of the photo so each of the sub-goals is assessed separately.

Since successful completion is the criterion, alternative methods to achieve the goal(s) (including errors that are successfully corrected) shall be included as positive results unless the means used have undesirable consequences, such as posing a risk to the well-being of the individual or causing damage to the product.

7.3.2.4 Establish criteria for sub-task outcomes

Goals and tasks can be decomposed into sub-goals and sub-tasks.

The sub-task outcomes that need to be achieved in order to use the product to achieve the goal(s) shall be identified to assist in planning the test and interpreting the results. The identification of sub-tasks could be based on information provided by the manufacturer, analysis of the product, and/or observation of a representative sample of potential users.

7.3.3 Specify the user group(s) to be used for testing

7.3.3.1 General

Identify the range of actual or potential users of the product. For consumer products, this should include both existing and new users, as appropriate. Different user groups may use the product to achieve different goals.

NOTE Some products are aimed at specific segments of the population (e.g. people with a particular hobby or a market segment related to income or age group) or may be in a family of related products for different market segments.

It is important to consider people who are towards the end of the range of user characteristics within the user population (see 7.3.3.4), especially for walk-up-and-use products and other products made available for use by the public. In particular are the following:

— users whose physical and cognitive capabilities are towards the ends of the range (for example, body dimensions, strength, biomechanical abilities, visual abilities, or auditory abilities), especially older users, and

— users whose cognitive characteristics are towards the ends of the range as a result of particular knowledge, experience, culture, literacy, or language.

A decision shall be made whether testing will include the whole range of users or whether a subset of these users is relevant to the aims of the testing.

Users with the widest possible range of characteristics and capabilities (within any subset) that are relevant to the testing should be tested (taking account of accessibility, see 7.3.3.4), and if any characteristics are expected to have a major impact on usability, each user group or contextual condition should be tested or analysed separately (see Annex C).

7.3.3.2 User characteristics

The characteristics that differentiate the people in the selected user groups and that are expected to have a significant effect on usability shall be identified, taking account of the user characteristics explained in ISO 20282-1, Clause 7, and summarized below (see C.2).

For those user groups to be included in the testing, the following information is needed:

— Which user characteristics could affect the usability of what is being tested?

— What range of characteristics exists in the intended or actual user group?

— What is the expected distribution of each relevant user characteristic?

a) Psychological and social characteristics

It is important to establish what cognitive characteristics or capabilities (such as knowledge and experience, cultural differences, and literacy and language) are likely to have a major effect on usability. This may be done by expert assessment, collection of actual usage data, or empirical studies to evaluate the effect of different user characteristics.

b) Physical characteristics

Identify whether the product requires particular physical characteristics (such as body dimensions, strength and biomechanical abilities, visual abilities, auditory abilities, or handedness). Either prior tests of users near the expected limits or the use of existing sources of data on the range of human characteristics may be used to establish whether the intended range of users can use the product. If the number of people who cannot use the product represents an unacceptable proportion of the intended users, a user-based test is inappropriate (see 7.4.2).

EXAMPLE 1 Existing data could be used to determine the size of text required for legibility by particular segments of the population or the height required to reach the controls of a vending machine.

c) Demographics

Demographics such as age, gender, and socioeconomic status provide a convenient way of grouping some physical, psychological, and social characteristics, but other differentiators (such as experience with particular types of products) can be more important.

EXAMPLE 2 The user population of a mobile phone are expected to be 80 % English speaking, 10 % French speaking, and 10 % who speak other languages. It is assumed that all the users will have prior experience in using mobile phones to make phone calls.

7.3.3.3 User groups

Decide whether the characteristics of the intended or actual user groups identified in 7.3.3.1 divide users into distinct categories that will have a major impact on usability (see C.2).

EXAMPLE 1 A global product is tested using separate groups for each nationality or language.

It is important to decide whether to test first-time users and/or people with previous experience. Even with a new product, people could have experience with similar products or a similar previous model so the type of relevant previous experience needs to be defined.

EXAMPLE 2 When testing a global positioning system (GPS), separate groups of users with and without experience with this kind of application are tested.

NOTE For setup and installation and for products used occasionally, it is appropriate to test first-time use. But for a product primarily used regularly, it could be better to test users after initial practice sessions.

7.3.3.4 Selecting user groups to assess accessibility

The selection of the user groups to be used for testing can provide information about the accessibility of a product.

If the product to be tested is walk up and use or a product made available for use by the public and when the product is intended to support people with the widest range of characteristics and capabilities whose physical or psychological characteristics are towards the extremes of the range within the general population, groups representing each extreme that is expected to have a significant impact on usability shall be selected for testing.

General demographic trends should be taken into account; for example, where the population is generally ageing, older people should be included unless the product specification actively excludes them.

Instead of testing a sample from the general population, test groups are selected from those that are most likely to find it difficult to achieve specified goals, in particular, goals relating to effectiveness. A positive result for effectiveness would usually also be valid for the members of the general population that are less likely to experience difficulties (see C.5).

EXAMPLE 1 Preliminary tests of an electronic voting machine have shown that it can be used successfully by people who are not at the extremes of the range of characteristics and capabilities that could influence their ability to use a voting machine. To provide additional assurance that it can be used by the full range of people who are eligible to vote, the tests are repeated with two subgroups, one with limited physical capabilities and the second a group over the age of 75 years with no computer skills.

EXAMPLE 2 Packaging is intended for use by people from the age of 8 years to 90 years. There is a particular interest in whether it can be opened by older people whose physical strength and sensory capabilities are likely to be less than those of younger users. Tests are carried out on a sample of users aged 65 years to 90 years. If the test group is successful in opening the packaging, it is predicted that the younger members of the user population will also be able to open it.

NOTE 1 The benefit of this approach is that if testing can be restricted to samples from the extremes of the population, the total number of participants can be reduced while still obtaining a relatively high degree of confidence normally associated with more participants.

NOTE 2 If the purpose of the test is to predict overall performance, tests on the full range of potential user groups would be needed.

Claims for the accessibility of a product shall be specific to the characteristics of the people involved in the testing. For example, having established that a product is usable by people with a limited level of acuity, this cannot be used to claim that it will be equally usable by people with colour deficiencies. Any statement about the accessibility of a product shall be accompanied by details of the user population of which the sample of participants is representative.

7.3.4 Specify environmental characteristics

The main environments in which the product is (or is intended to be) used for the selected users and tasks shall be identified. Those attributes of the physical, social, or technical environment that are

likely to have significant effects on the effectiveness, efficiency, or satisfaction with which goals can be achieved shall be specified.

EXAMPLE The camera could be used at night where there might be no other illumination. The temperature ranges from –5 °C to 40 °C.

7.4 Check that the product is capable of achieving the intended goals for the intended users

7.4.1 Check that the product is technically capable of meeting the criteria for successful goal achievement

For each goal to be tested, a check shall be made that the product is capable of producing results that can adequately satisfy the needs of the users. Simple goals can be checked by inspection (e.g. that an on/off switch performs correctly). For more complex goals (e.g. that an alarm is loud enough or that an oven heats to the indicated temperature), the check shall be based on one of the following:

a) Documentation provided by the manufacturer states that the performance of the product meets established technical standards.

b) Carry out technical tests based on International Standards or other published criteria for what constitutes acceptable quality (e.g. that the controls in a lift are within the reach envelope for people in wheelchairs).

c) Carry out expert assessment of the technical quality, where possible, based on International Standards or other published criteria for technical results that fall within an acceptable range.

d) For installation and setup, an expert can check that the product can be installed and set up in way that would adequately meet user goals.

If none of the above sources of evidence is available, testing achievement of the goal is outside the scope of this part of ISO/TS 20282 and the absence of evidence for support for this goal shall be included in the report.

If as a result of the checks above, any goals cannot be achieved adequately with the product, they shall be excluded from the testing of the overall usability of that product.

7.4.2 Check that the product is compatible with the characteristics of the intended users

A check shall be made that the intended user groups can, in principle, operate the product, e.g. that the controls are within physical reach and/or that users have essential prerequisite knowledge.

NOTE 1 Some user groups, such as older or disabled users, might use alternative modes of interaction.

Where inspection of the product indicates that successful attainment of any of the goal(s) identified in 7.3.2.2 will not be possible for some members of an intended user group, information about the absence of support for this goal shall be included in the report.

NOTE 2 Alternatively, it might be appropriate to reconsider the design of the testing.

NOTE 3 ISO/TR 22411 contains ergonomics data and knowledge about human abilities for older persons and persons with disabilities on which such inspections can be based.

EXAMPLE The intended user population for a ticket machine includes a specific subgroup of intended users: children from the age of 8 years with a criterion for successful goal achievement of 95 %. It can be shown that users whose upward arm reach is less than 1,5 m will not be able to operate the machine to obtain tickets. The relevant anthropometric data for the age group enable those responsible for the evaluation to calculate that 50 % of this group will not be able to obtain tickets from the machine.

7.5 Design and perform the test

7.5.1 General

One test session can include tests on different products, users groups, and/or goals and may involve more than one task. The measures described in the remainder of 7.5 are those that need to be identified for each task.

7.5.2 Identify the relevant measures

7.5.2.1 General

The required measures (in bold) and examples of optional attributes that can be measured are given in Table 1.

Table 1 — Examples of measures of attributes of usability

Attribute	*Objective measures*	*Subjective measures (questionnaires)*
Effectiveness	**Success rate**	Perceived success
Efficiency: *a) resources utilized*	**Task time,** Cost of performing the task	Perceived time, Perceived cost
b) demand on human resources	Cognitive or physical effort Consequences of mental workload (e.g. secondary task performance)	Perceived mental effort Perceived physical workload
Satisfaction	Psychophysiological measures [e.g. galvanic skin response (GSR), pupil dilation] — not well established as they are unreliable One can measure potential consequences of satisfaction, e.g.: — reuse of the product or service — purchase of the product or service — change in brand	**Satisfaction with the task achievement or with the product** Satisfaction with task time Measures of pleasure Measures of user experience Measures of trust Measures of comfort Propensity to recommend

The relative importance of different measures will depend on the type of goal(s) and the context of use.

7.5.2.2 Effectiveness

Effectiveness shall be measured by the **success rate**: the percentage of users who successfully achieve each goal of use of the product (see 7.3.2.3).

NOTE If more time is taken than the maximum permitted task time, then the task **is** categorized as unsuccessful (see 7.5.2.3).

The participants' perception of success (**perceived success**) should also be assessed for tasks where there could be any uncertainty.

7.5.2.3 Efficiency

a) Resource utilization

The time taken to achieve the goal shall be measured.

To enable task timing, the criteria for task completion shall be specified. This should include a statement by the users that they believe that they have completed the task. Task time measurements shall only include time to perform the task itself, not the time to read and comprehend the task instructions. Timing should start after the user has read the instructions and at the point that information on the interface is presented to the user. The test task instructions shall be available to the participant for review during the task performance.

A maximum amount of time allowed for successful goal achievement shall be established. This time should be sufficient to allow test persons to try different ways of achieving the goal before they give up. The expected time to complete the task should be estimated from pre-tests and the maximum time allowed to users before they are categorized as having failed should be at least three times the expected time. Task time measurement should stop if more time is taken than this maximum time and the task shall then be categorized as unsuccessful.

Cost of performing the task: Although time is the most commonly measured resource, in some circumstances, other resources that cost money can be important (such as the cost of a mobile phone call).

Perceived time, perceived cost: It could be important to know the users' impression of the time or cost.

b) Demand on human resources

Cognitive or physical effort shall be measured when the task makes a significant demand on cognitive or physical resources.

Physical effort: For physical controls, there will be an optimal level of physical effort and a maximum acceptable physical effort that could be measured objectively or subjectively.

Mental effort: In addition to maximum acceptable mental effort, there can also be a minimum for some tasks (for example, to maintain interest). Mental effort is most commonly measured by a questionnaire.

7.5.2.4 Satisfaction

Overall satisfaction for all the goals included in the test shall be measured immediately at the end of testing using the Smiley scale in E.1. SUS can also be used.

NOTE 1 The Smiley scale has the advantage that it is simple and fast to administer and widely used in consumer research.

NOTE 2 The SUS scale is widely used in usability research and, as it has 10 questions, is likely to be more reliable. The distribution of results for different types of products is available.[24]

Measures of satisfaction can include overall satisfaction with the product, satisfaction with specific features of the product, satisfaction with the adequacy of goal achievement, and enjoyment of use.

When more than one task is tested, if it is appropriate, satisfaction should be measured for each task as well as for overall usage (see E.2).

NOTE 3 Measuring satisfaction after the whole test will more accurately capture the users' overall impression, while measuring satisfaction after each task (for example, with the Smiley scale) will produce a more accurate assessment of the users' satisfaction with each goal or task.

When other aspects of satisfaction such as the users' experience, pleasure, or trust are important, more specialized scales can be used as well. When a published and validated questionnaire exists, this should be used because it will produce more reliable results than an *ad hoc* questionnaire.

NOTE 4 The consequences of satisfaction can include influencing behaviours such as purchase or reuse of a product or brand loyalty.

7.5.3 Determine the number of participants required for the desired confidence level

To reduce the effect of any sampling error, a minimum of 10 representative participants shall be used for each user group tested (see C.3).

The required statistical confidence level (confidence that the result has not occurred by chance) shall be identified. The degree of confidence needed will depend on the type of test and the purpose. The higher the required success rate and the higher the confidence, the more users will be required.

NOTE In scientific work, it is usual to require at least 95 % confidence to draw a conclusion, but for some business purposes, a lower degree of confidence (such as 80 %) could be sufficient[21] if the 20 % risk of an incorrect conclusion can be accepted.

EXAMPLE 1 The International Standard for child-resistant containers, ISO 8317, requires that 75 % of children are unable to open a container, with 95 % confidence.

EXAMPLE 2 Table D.2 shows that to be 95 % confident of an 80 % success rate, with a sample of 14 users, all of them would have to be successful or 20 out of a sample of 21 users would have to be successful. Table D.1 shows that if 80 % confidence is sufficient, samples as small as 4 could give 80 % confidence of an 80 % success rate.

7.5.4 Recruit a sample of users

A sample (or samples) representing the intended or actual user group(s) shall be recruited (from the user population identified in 7.3.3) to take part in the test. The sample(s) of users shall be selected to model the distribution of relevant user characteristics within the specified user group(s) (see C.4). The sample(s) should have a distribution of values or attributes of the relevant user characteristics that is as close as possible to the intended or actual user group(s) identified in 7.3.3.

NOTE 1 Many general, physical, and psychological characteristics (such as strength, visual acuity, and cognitive capabilities) can be adequately sampled through demographic variables such as age and education. However, this approach will not account for factors such as specific disabilities, relevant previous experience, specialized knowledge or skills, and culture. Such groups can be selected based on the product and the interface requirements for successfully interacting with the product and tested separately (see C.1).

EXAMPLE An alarm clock requires manual dexterity to operate the controls and a reading age of 12 years to understand the instructions. These relevant user characteristics are sampled by recruiting users with a representative distribution of age and educational attainment. Only 5 % of purchasers are expected to have previous experience with this type of alarm clock, and as the task of setting the alarm is expected to be much more difficult for those who do not have this experience, only people without previous experience will be tested.

Test participants shall not be from the testing or manufacturing organization unless the product is to be used within the organization, in which case test participants shall not be from the development department. Efforts should also be made to recruit outside family and friends as they may be biased.

NOTE 2 Participants that are familiar with the technology domain and/or the design and development process possess knowledge and skills that representative users do not. Including them in a study could compromise the study's validity.

7.5.5 Configure the product

The version of the product that is tested shall be recorded. Evaluation shall be conducted with the product being used in conjunction with any other equipment or materials that are normally involved in its use.

Brand can have a major influence on satisfaction so if it is appropriate to reduce the influence of brand on results, the brand of the product should be hidden.

The product should be installed in the appropriate configuration (usually the default configuration as delivered), or for installation and setup, it shall be provided as it would typically be received by the user.

When the test commences, the product should be set in a typical configuration that would be experienced by users.

EXAMPLE Using a mobile phone with its default settings or using a TV recorder placed under a TV. For installation, a mobile phone is provided in a box with associated documentation. An oven is provided professionally installed in a physical housing connected to gas or electricity, but without configuration of the clock or other controls.

Any instructions that are provided with the product shall be made available to the users, unless the purpose of the test is to test the usability or accessibility without instructions.

7.5.6 Specify the test scenarios and goal criteria

A test scenario shall be specified for each goal that is to be tested.

A test scenario is an instance of the context of evaluation describing a situation in which the participant is requested to achieve specific goal(s) in a particular environment.

The test scenario shall include a precise statement of the goal to be attained (see 7.3.2).

EXAMPLE 1 Make a phone call to 01509 123456 using this telephone.

EXAMPLE 2 Set up a mobile phone for first use.

For each test goal, the criteria for goal attainment shall be identified (see 7.3.2.3), together with a way of establishing whether it has been achieved.

It is sometimes possible to set criteria for acceptable task time,[27] for example, based on how long the user would expect it to take to achieve the same goal using a different product. The maximum acceptable time shall be specified for each task (see 7.5.2.3).

If there is more than one goal to be tested, it will be necessary to decide the order in which the tasks are to be undertaken. In general, if there is a normal sequence in which tasks would be undertaken, this shall be the order for testing. If no normal sequence exists, testing in one fixed sequence shall be avoided in order to avoid arbitrary order effects. In this case, the order should be systematically varied or assigned randomly for each person tested.

7.5.7 Establish the test procedure

a) A written script shall be prepared for each task goal, containing instructions that are presented to each user. The script shall include a description of the scenario within which testing is taking place (e.g. taking a photo at a family gathering) and shall state the particular goal and the conditions which apply to it. No hints shall be included on how to achieve the goal or which features to use. The script should contain all possible information that the user needs to perform the task. Participants shall not be required to read any product instructions that are provided (see 7.5.5).

b) Users shall not be given any hints or assistance from the personnel conducting the test, but depending on the purpose of the test, the instructions that come with the product may be provided.

c) Those conducting the test shall note the time taken to achieve each goal. If the user reaches the maximum amount of time allowed without attaining the goal, they shall be asked to stop, and if appropriate, move on to the next goal.

d) The nature of any errors or other difficulties encountered by users may be recorded in order to identify usability problems that can be provided as feedback to design. This could include which sub-goals were achieved or not achieved, and why.

7.5.8 Create the test environment

a) Testing shall be carried out in an environment that resembles as closely as possible the environment in which the product would be used. If a usability laboratory is used, the normal conditions of use shall be simulated. If tested in the field, any variations in environmental conditions that might have effects on the level of usability achieved should be recorded.

   If goal attainment while using the product could be adversely affected by particular environmental conditions, testing could be carried out in those conditions in order to be ensured of high levels of usability in all situations.

   EXAMPLE 1 Normally used in cold weather by people wearing gloves so tested in a cold environment.

b) If the product is normally used/installed in a stressful environment, this should either be reproduced in a test environment or a field test used.

EXAMPLE 2 Testing a route guidance system in heavy traffic conditions.

c) Provide any materials needed to use/install the product.

d) Access may be provided to a helpline or Internet help, if provided for the product.

EXAMPLE 3 Ironing board, notes for a machine, bread for a toaster, clothes for a washing machine.

7.5.9 Pilot test

When testing a product for the first time, having one or two people complete the usability test can reveal obvious flaws with the test design or with the product prior to the full test. A pilot test can result in the revision of task scenarios to reduce ambiguity and avoid unexpected problems.

7.5.10 Assess whether the criteria for goal achievement have been met by individual users

For each user, success for each goal tested shall be judged by assessing the outcome to find out whether the success criteria have been met (see 7.3.2.3). One of the following methods shall be used for making the assessment:

a) judgement by the test administrator of success by inspection of the outcome, when this can be done in a valid and reliable way;

b) judgement of success based on a sequence of user actions that are known to reliably result in an acceptable outcome (see 5.3);

c) use of established technical criteria to assess the quality of the results.

8 Results

8.1 General

All measures shall be reported with confidence intervals (see Annex D).

If claiming that a requirement has been met, the statistical level of significance shall be stated.

If making a comparison between products, the statistical significance of any differences between products or other conditions shall be stated.

8.2 Goals

If measuring the usability of the whole product, and if all the main goals of the product have been tested with the same results (pass or fail), the report may make an overall statement about the usability of a product.

Otherwise, if more than one goal is tested, the test results shall specify which goals were tested and why, and results shall be given separately for each goal. When comparing products, either the main goals shall be identified and used as a basis of testing or the goals used for testing shall be stated with an explanation of why they were selected.

8.3 Effectiveness

The result(s) reported shall include the percentage of users successfully achieving each goal (success rate), together with the confidence intervals, as specified in Annex D.

EXAMPLE 11 out of 12 users achieved the main task goals of use of the product. It is estimated with 80 % confidence that at least 82,1 % of the intended users will be successful.

8.4 Efficiency

The result(s) reported shall include the geometric mean time[29] taken by users who successfully achieved the specified goal(s), together with the confidence intervals.

NOTE The geometric mean is the mean of log-transformed data, transformed back to the original scale.

EXAMPLE The median time taken by those users who successfully bought the correct ticket was 20 s (with a 95 % probability that the true value is between 15 s and 25 s).

Additional measures of efficiency (see 7.5.2.3) may also be reported.

8.5 Satisfaction

The result reported shall include the mean[30] Smiley scores (see Annex E) together with the confidence intervals.

9 Report

A full report of the results shall be prepared using the format specified in Annex F.

NOTE A full report provides the information necessary to make future comparable tests.

If a short version of the report is provided, it should include the information in the executive summary in F.2.

If all the main goals of the product have been tested with similar results, the statement may include an overall statement about the usability of a product.

Any statement of results shall include the following information:

a) goals not supported (see 7.4);

b) if the main goals are tested, an explanation of how the main goals were identified (see 7.3.2.2);

c) if the main goals are not being tested, the reasons for selecting the goals being tested (see 7.3.2.2).

Annex A
(informative)

User goals

Table A.1 gives examples of the main user goals for selected products.

NOTE It might not be sufficient to only test these main goals if there are other important goals.

Table A.1 — Examples of different products and the main associated user goal(s)

Product	Main goal(s)	Example sub-tasks	Installation issues
Alarm clock	Get woken at a certain time	Adjust time, set alarm time, switch alarm on, switch alarm off	Insertion of batteries is an installation and servicing task
Blood pressure measuring device	Obtain a measurement of blood pressure	Put on, start measurement, take reading, take off	Insertion of batteries is an installation and servicing task
Camcorder	Record a video and play it back	Insert recording media, take video, stop, play movie	
Cash dispenser (ATM)	Obtain cash using debit card	Insert card, insert PIN, key in amount, take card, take money, recover card	Installation is not the responsibility of the consumer
Cassette recorder	Record and play auditory information	Insert new cassette, control levels, make recording, stop recording, rewind tape, play tape, adjust volume, remove tape	Setting up with a stereo is an installation task
CD player	Listen to a CD	Insert CD, play CD, select track, stop CD, eject CD	
Dishwasher	Obtain clean dishes	Insert dishes, insert detergent(s), select programme and start machine, stop machine, get out dishes	
Dryer	Dry clothes	Insert clothes, select programme, switch off machine, remove clothes	Cleaning the filter is a servicing task
Electric heater	Heat an area	Switch on, select heat, switch off	
Electric hob	Cook food	Switch on selected hob to specific temperature, adjust temperature during cooking, turn off hob	
Electric iron	Remove creases from clothes	Set iron to desired temperature, use iron to remove creases, turn off iron	
Electric mixer	Mix different types of food	Unravel cord, plug in, insert food and mixer tool, switch on, mix, switch off, get food, clean, pack up	
Electric plate	To have hot food	Switch on selected plate with selected heat	
Electric oven	Cook food	Heat oven to desired temperature	
Fax machine	Send page(s) to a specified fax number	Insert page, dial and send fax to number, get back paper	Connection to a phone line and switching on are installation tasks

Table A.1 *(continued)*

Product	Main goal(s)	Example sub-tasks	Installation issues
Fire extinguisher for kitchen	Extinguish burning liquid	Take extinguisher from place, activate extinguisher	
Gas hob	Heat up food	Switch on selected plate, ignite gas, adjust heat	
Light switch	Illuminate a specific location	Switch on selected room's light, switch off selected room's light	
Locker	Put something in a safe place and retrieve it later	Insert money, open locker, store item, close locker, keep key or code, open locker, retrieve item	
Microwave	Heat food to an appropriate temperature	Insert food, select power and time, start, take out food	
Mobile telephone	a) Take a call b) Make a call	Activate phone, activate number (dial)	
Packaging	Obtain access to the contents	Make opening, remove surplus packaging	
Parking ticket vending machine	Comply with the parking rules in a specific location	Insert payment, remove and keep ticket, display ticket in car	
Photo camera	Take pictures for later review	Insert film or storage media, switch on, take aim, take picture, take steps to review picture	Insertion of batteries is an installation and servicing task
Product vending machine	Buy a product	Select a product, insert payment, get product	
Public telephone	Make a call to a specific number	Activate phone (lift receiver), make payment, activate number (dial)	
Public coffee vending machine	Get a single cup of coffee or a particular type	Select type, insert payment, get coffee	
Radio	Listen to a particular radio station	Switch on, select station, adjust volume, switch off	
Refrigerator	Store food at an appropriate temperature	Insert food, select appropriate temperature, take out food	
Room telephone	a) Take a call b) Make a call	Activate phone (lift receiver), activate number (dial)	
Satellite set-top box	Watch selected TV channels	Switch on, select station, switch off	Setting up is an installation task
Telephone answering machine	Record and play incoming calls	Play received calls, play a specific received call, delete all calls	Recording messages belongs to setting up task
Television with remote control	Watch selected TV channels	Switch on TV, select programme and volume with remote control, switch off TV	Setting up of channels is an installation task
Toaster	Get bread toasted to appropriate brownness	Insert bread, select brownness, switch on, take out bread	
Vacuum cleaner	Clean an area of floor	Unwind cord, plug in, switch on, select strength, move brush, switch off, unplug, ravel cord	Replacing dust bins is a service task

Table A.1 *(continued)*

Product	Main goal(s)	Example sub-tasks	Installation issues
Personal video recorder (PVR)	Record and play videos	Switch on, locate programme guide, select programme to be recorded, locate list of recordings, select programme, play programme, switch off	Setting up with a TV is an installation task
Transport ticket vending machine	Buy an appropriate ticket for a journey	Select one adult, one-way fare to destination, insert payment, get ticket	
Washing machine	Clean some clothes using an appropriate programme	Insert clothing, insert detergent, select appropriate programme, start machine, take out clothing after machine stops, switch off machine	
Water tap	Obtain water of required temperature	Turn on, adjust temperature, turn off	

Annex B
(informative)

Background to the method

B.1 Difference between formative and summative usability test methods

Summative test methods are used to measure how usable a product is. This part of ISO/TR 20282 contains a summative test method that may be used, for example, to estimate the percentage of people that can successfully achieve their goals in actual use. Formative test methods are intended to be used to understand the strengths and weaknesses of a product in order to improve its usability. Formative user-based usability test methods typically employ fewer users and involve the user thinking aloud so the methods do not give accurate estimates of effectiveness, efficiency, or satisfaction.

B.2 Choice of measures

To obtain reliable measures of usability, a summative evaluation method consistent with ISO 9241-11 is needed. This contrasts with typical formative evaluation methods that use a small number of participants to identify usability problems.

Effectiveness is defined in ISO 9241-11 in terms of accuracy and completeness, but for the simple tasks within the scope of this part of ISO/TR 20282, only success rate is used.

NOTE Accuracy is quite difficult to measure; see, for example, ISO/IEC 25062.

Efficiency is defined in ISO 9241-11 in terms of the resources expended in relation to the accuracy and completeness. For practical purposes, time is the most important resource (and is more easily interpreted on its own rather than as a ratio).

Satisfaction is defined in ISO 9241-11 as freedom from discomfort and positive attitudes towards the use of the product. For usability, the most important component of satisfaction is a positive attitude.

B.3 Number of users

While small numbers of users are sometimes sufficient to identify a large proportion of usability problems, larger numbers are required to confidently estimate the population mean for success rate (see C.3). Larger numbers also make it easier to obtain a representative sample of users.

Annex C
(normative)

Recruiting a representative sample of users

C.1 Decide whether to divide users into groups

Decide whether the characteristics of the intended or actual user groups identified in 7.3.3 divide users into distinct categories that will have a major impact on usability.

If there is more than one distinct user group, decide whether to test all or only some user groups, depending on the purpose of the test.

EXAMPLE 1 If the language spoken will have a major impact on usability, each language group is tested separately.

EXAMPLE 2 If previous knowledge or experience will have a major impact on usability, each group is tested separately.

EXAMPLE 3 Initial testing might just be to validate usability for one predominant user group.

EXAMPLE 4 If usability is known to be acceptable for the predominant user groups, only user groups that could have more difficulty (such as the elderly or people with disabilities) might be tested.

EXAMPLE 5 To assess whether usability and accessibility for users with the widest possible range of characteristics has been achieved, user groups drawn from extremes of the range of variance of the characteristics would be tested.

EXAMPLE 6 If information is required about usability for the whole population, one representative sample drawn from the population could be tested. To maximize statistical validity, use as large a sample size as possible. But it will not be possible to draw any statistical inferences about the usability for any constituent groups that represent only a small percentage of the population.

If after testing, inspection of the test data suggests that there could be larger than expected differences between subgroups within one user group, the product may be tested again using new samples of users for each subgroup.

It is important to decide whether to test first-time users and/or people with previous experience. Even with a new product, people could have experience with similar products or a similar previous model so the type of relevant previous experience may need to be defined.

EXAMPLE 7 Users of a mobile phone are divided into groups determined by their expertise in setting up this type of mobile phone:

a) have never set up a mobile phone;

b) have set up a mobile phone of a different brand;

c) have previously set up a similar mobile phone.

Each group will be tested separately, as they are expected to have different results for usability.

Only one group might be tested, for example, the most frequent users or a critical user group.

EXAMPLE 8 A hotel alarm clock is expected to be difficult to set for the first time, but easy to set on subsequent occasions. Only those who have not used the alarm clock before are tested.

C.2 Choose sample size

To reduce the effect of any sampling error, a minimum of 10 representative participants shall be used for each user group.

The larger the sample size, the more accurate the prediction (provided the sample is representative of the user population). If the results obtained from the initial sample of users are not sufficiently accurate, additional participants can be tested.

EXAMPLE If the measured success rate is 90 %:

— For 9 successful out of 10 users, the minimum population success rate with 80 % confidence is 79 %.

— For 18 successful out of 20 users, the minimum population success rate with 80 % confidence is 81 %.

— For 45 successful out of 50 users, the minimum population success rate with 80 % confidence is 86 %.

— For 90 successful out of 100 users, the minimum population success rate with 80 % confidence is 87 %.

C.3 Recruit a representative sample of users

User groups are differentiated by the characteristics of the users, tasks, or environments that are expected to have a major impact on usability. Within each user group, there will be many other characteristics (such as age or gender) that could possibly have an impact on usability. A sample that is representative of these characteristics should be recruited to take part in the test. Stratified sampling may be used to obtain a sample that represents the major relevant characteristics. If tested in the field, a representative sample could be constructed by interviewing normal users to select those to be included in the sample.

Many general physical and psychological characteristics (such as strength, visual acuity, and cognitive capabilities) that can affect usability can be indirectly sampled through appropriate demographic variables such as age, sex, and educational attainment. Users should be selected to be representative of these demographic variables and any other significant characteristics that will not be adequately sampled by the demographic variables.

EXAMPLE 1 In a multi-ethnic population, users are selected to have a range of cultural backgrounds that is representative.

EXAMPLE 2 When testing a consumer product, users are selected to be representative of their experience with different brands.

EXAMPLE 3 An alarm clock requires manual dexterity to operate the controls and good visual acuity to read the instructions. These relevant user characteristics are sampled by recruiting users with a representative distribution of age, capabilities, and educational attainment. Because 5 % of purchasers are expected to have previous experience with this type of alarm clock and the task of setting the alarm is expected to be much more difficult for those who do not have this experience, only people without previous experience will be tested.

Stratified sampling (see ISO 15535) should be used to distribute users across as many permutations of demographic characteristics that are not expected to have a sizable effect on usability as is practical. The preferred combinations of user characteristics can be laid out in a table.

EXAMPLE 4 Table C.1 shows a sample of 12 participants in a user group who have been selected so that they include equal representation of the attributes: age in the range of 18 years to 65 years, gender, and level of education.

Table C.1 — Example of a stratified sample

Age/Sex	Highest level of education				Total
	School		Higher education		
	Male	Female	Male	Female	
18–33	1	1	1	1	4
34–58	1	1	1	1	4
59–65	1	1	1	1	4
Total	6		6		12

C.4 Minor user groups

If it is not possible to include minor user groups (such as a specific nationality or specific disability) because they constitute a very small percentage of the intended user population within a representative sample, additional tests may be carried out with separate samples for each group to determine whether members of these groups are able to use the product.

The composition of samples for testing shall not include members of user groups with characteristics that are expected to have significant effects on usability. If users with these characteristics are part of the intended user population, a representative sample should be tested as a separate user group.

NOTE If usability is known to be acceptable for most of the user population, testing could be restricted to specific groups that are expected to have more difficulty.

C.5 Frequency of participation

If participants are recruited from a database of volunteers, there should be a limit on how frequently people are permitted to participate. This is to avoid creating “professional participants”. This could be based on the guidelines used in marketing research for focus groups (typically restricting participation to twice a year).

C.6 Normal usage data

Data may be obtained from a walk-up-and-use product in normal use, either by observation (if this is acceptable) or from data collected automatically from a machine that has been instrumented to log user actions and responses.

NOTE Data from logs could provide good estimates of task time for completed tasks. The success rate data would be of less value, as it would only record information on people who initiated but did not complete their task. It would not include people who decided not to start the task.

Annex D
(normative)

Confidence intervals

D.1 Effectiveness

Effectiveness is measured as the percentage of users who successfully achieve a goal of use of a product. The confidence intervals for effectiveness shall be calculated using the exact binomial statistic (see Bibliography). To establish a high success rate with a high degree of confidence requires a large sample of users. Table D.1 gives examples of statistical calculations (using the Adjusted Wald method) of the maximum number of unsuccessful users required to estimate population success rates greater than 75 % and 80 % with 80 % and 95 % confidence using one-tailed test (as the hypothesis is only rejected for values that are significantly lower).

NOTE 1 The Adjusted Wald method is used to calculate confidence intervals as this method provides population values for small samples.[25] [28] [31]

EXAMPLE If 9 out of 10 users are successful, then there is a 95 % probability that the success rate is at least 63 % and an 80 % probability that the success rate is at least 78 %.

NOTE 2 These statistics are only valid for random samples with a distribution of test results that approximates to normal. It is important that the users are selected randomly, and when stratifying a sample, it is important that the users are selected randomly in each category (see Annex C).

NOTE 3 A 95 % level of confidence is conventional in scientific work and where definitive claims are being made; however, an 80 % level of confidence may be sufficient for business decisions.

Table D.2 shows that the maximum number of unsuccessful users required to estimate a population success rate greater than 80 % with 95 % confidence is

— 0 users out of 14 to 20 users tested,

— 1 user out of 21 to 28,

— 2 users out of 30,

— 3 users out of 40,

— 5 users out of 50, and

— 13 users out of 100.

The maximum number of unsuccessful users required to estimate a population success rate greater than 75 % with 95 % confidence is

— 0 users out of 10 to 16 users tested,

— 1 user out of 17 to 21,

— 2 users out of 22 to 27,

— 3 users out of 30,

— 5 users out of 40,

— 7 users out of 50, and

— 17 users out of 100.

NOTE 4 Examples of other estimates of population success rates are given in Tables D.1 and D.2. In these tables:

Column 1: Number of participants in the test.

Column 2: The maximum number of participants in the test that can be unsuccessful when the purpose of the test is to show with 80 % confidence that the population success rate is greater than 75 %.

Column 3: For each indicated number of unsuccessful participants, the estimated minimum percentage of people in the population who can use the product successfully, calculated with 80 % probability.

Column 4: The maximum number of participants in the test that can be unsuccessful when the purpose of the test is to show with 80 % confidence that the population success rate is greater than 80 %.

Column 5: For each indicated number of unsuccessful participants, the estimated minimum percentage of people in the population who can use the product successfully, calculated with 80 % probability.

Table D.1 — Examples of statistical calculations of 80 % confidence intervals

1	2	3	4	5
Estimated population success rate	75 %		80 %	
No. of users	Max. no. of unsuccessful users	Estimated minimum population success rate	Max. no. of unsuccessful users	Estimated minimum population success rate
3	0	77,6 %		
4	0	82,2 %	0	82,2 %
5	0	85,3 %	0	85,3 %
6	0	87,5 %	0	87,5 %
7	0	89,1 %	0	89,1 %
8	0	90,3 %	0	90,3 %
9	1	76,7 %	0	91,3 %
10	1	78,8 %	0	92,1 %
11	1	80,6 %	1	80,6 %
12	1	82,1 %	1	82,1 %
13	1	83,3 %	1	83,3 %
14	2	75,9 %	1	84,4 %
15	2	77,4 %	1	85,4 %
16	2	78,7 %	1	86,3 %
17	2	79,9 %	1	87,0 %
18	2	81,0 %	2	81,0 %
19	3	75,9 %	2	81,9 %
20	3	77,0 %	2	82,8 %
21	3	78,0 %	2	83,5 %
22	3	79,0 %	2	84,2 %
23	3	79,8 %	2	84,9 %
24	4	75,9 %	3	80,6 %
25	4	76,8 %	3	81,4 %
26	4	77,7 %	3	82,0 %

Table D.1 *(continued)*

1	2	3	4	5
Estimated population success rate	75 %		80 %	
No. of users	Max. no. of unsuccessful users	Estimated minimum population success rate	Max. no. of unsuccessful users	Estimated minimum population success rate
27	4	78,5 %	3	82,7 %
28	5	75,2 %	3	83,3 %
29	5	76,0 %	3	83,8 %
30	5	76,8 %	4	80,5 %
40	7	76,9 %	5	82,4 %
50	9	77,0 %	7	81,3 %
100	21	75,4 %	16	80,7 %

Table D.2 — Examples of statistical calculations of 95 % confidence intervals

1	2	3	4	5
Estimated population success rate	75 %		80 %	
No. of users	Max. no. of unsuccessful users	Estimated minimum population success rate	Max. no. of unsuccessful users	Estimated minimum population success rate
10	0	75,1 %		
11	0	76,9 %		
12	0	78,4 %		
13	0	79,7 %		
14	0	80,9 %	0	80,9 %
15	0	82,0 %	0	82,0 %
16	0	82,9 %	0	82,9 %
17	1	76,0 %	0	83,8 %
18	1	77,2 %	0	84,5 %
19	1	78,2 %	0	85,2 %
20	1	79,1 %	0	85,9 %
21	1	80,0 %	1	80,0 %
22	2	75,1 %	1	80,8 %
23	2	76,0 %	1	81,5 %
24	2	76,9 %	1	82,2 %
25	2	77,7 %	1	82,8 %
26	2	78,5 %	1	83,4 %
27	2	79,2 %	1	83,9 %
28	3	75,5 %	1	84,4 %

Table D.2 *(continued)*

1	2	3	4	5
Estimated population success rate	75 %		80 %	
No. of users	Max. no. of unsuccessful users	Estimated minimum population success rate	Max. no. of unsuccessful users	Estimated minimum population success rate
29	3	76,2 %	2	80,4 %
30	3	76,9 %	2	81,0 %
40	5	76,2 %	3	82,2 %
50	7	75,9 %	5	80,6 %
100	17	75,9 %	13	80,4 %

D.2 Efficiency

The confidence intervals for efficiency shall be calculated using the standard error of the mean for task time. If the task time data are not normal, it should be transformed to a normal distribution, for example, using a logarithmic transformation to calculate the geometric mean before calculating the standard error of the mean.

NOTE 1 The standard error of the mean is the sample standard deviation divided by the square root of the sample size.

NOTE 2 The size of the confidence interval will depend on the size of the sample and the variance in the population tested. The variance will generally decrease with increased sample size and will generally be larger for more diverse populations.

EXAMPLE The standard error of the mean is calculated after a logarithmic transformation and is equivalent to 4,2 s. The median task time is 22,2 s so the results could be reported as a median of 22,2 s, with a 95 % probability that the true value is between 18,0 s and 26,4 s.

D.3 Satisfaction

The confidence intervals for satisfaction shall be calculated using the standard error of the mean.

Annex E
(normative)

Examples of satisfaction scales

E.1 Smiley scale

The Smiley scale, which was first brought up by Kunin (1955),[20] has been widely used in psychology to measure emotional aspects of satisfaction and has also been applied in industrial usability tests. A five-point version of the scale has recently received factor analytical validation and has been positively tested for equidistance by Ruth Jäger (2004).[18]

The circles shall be 10,5 mm in diameter and the scale should be used as is stretched over 69 mm between the centres of the extremes.

For users with visual disabilities, these sizes may be increased proportionately.

To use the scale, the faces presented in Figure E.1 shall be used as they are, as the convexness and concaveness of every Smiley's mouth has been carefully selected for equidistance.

"Please tick the face that corresponds most closely with your degree of satisfaction with…"

Figure E.1 — Smiley scale

The Smileys correspond to the values −2, −1, 0, +1, and +2. No intermediate measures are allowed. Group values may be given as an arithmetic average.

E.2 Satisfaction questionnaires

Satisfaction questionnaires are widely used and validated questionnaires with multiple questions can measure different aspects of satisfaction to produce more reliable results[15] than a single-item questionnaire, but single-item questionnaires such as the Smiley scale, SMEQ,[32] or a Likert scale[26] are more convenient to administer after each task.

NOTE Although satisfaction questionnaires produce ordinal data, applied statisticians have found that parametric statistics produce meaningful results when analysing satisfaction questionnaires.[30]

ISO/IEC 25010 distinguishes the following aspects of satisfaction that could potentially be measured:

— Usefulness: the degree to which the stakeholder is satisfied with their perceived achievement of pragmatic goals, including acceptable perceived results of use and consequences of use (using, for example, SUS[15]).

— Trust: the degree to which the stakeholder is satisfied that the product will behave as intended (using, for example, the System Trust Scale[19]).

— Pleasure: the degree to which the user obtains pleasure from fulfilling their personal needs. Personal needs can include to acquire new knowledge and skills, to communicate personal identity, and to provoke pleasant memories. If pleasure is an important issue, a questionnaire such as AttrakDiff[17] could be used.

— Comfort: the degree to which the user is satisfied with physical comfort (using a questionnaire such as CCBaseline[22]).

Annex F
(normative)

Format for test reports

NOTE This format is based on the Common Industry Format for usability test reports (ISO/IEC 25062).

F.1 Title page

The following information shall be provided:

a) title: ISO/TS 20282-2 *Test report for usability*;

b) product and version that was tested;

c) when the test was conducted;

d) where the test was conducted;

e) date the report was prepared;

f) test organization name;

g) test organization contact person and contact details.

F.2 Executive summary

This section provides a high-level overview of the test. The intent of this section is to provide a stand-alone summary.

A high-level overview of the test shall be provided, suitable for reading by a wider audience, which includes

a) name and description of the product,

b) user group(s), goal(s), and environment(s) that have been tested,

c) any instructions included with the product that were provided to the users and whether they were used, and

d) results for each goal expressed as mean scores or other suitable measure of central tendency with confidence intervals.

F.3 Product and expected context of use

F.3.1 Full product description

The following information shall be provided:

a) formal product name, model, and version;

b) parts or functions of the product that were evaluated.

F.3.2 Expected context of use

The following information shall be provided:

a) goals tested and how these were selected;

b) intended or actual user groups of the product and how these were identified;

c) for those user groups included in the testing, the following information is needed:

   — user characteristics that could affect the usability of the product;

   — range of characteristics that exists in the intended or actual user group;

   — expected distribution of each relevant user characteristic;

   — suitability of product for older users or any other groups with special needs;

d) main environments in which the product is (or is intended to be) used.

F.4 Method

F.4.1 General

Sufficient information shall be provided to allow an independent tester to replicate the procedure used in testing.

F.4.2 Participants

The following information about participants shall be provided:

a) any physical, sensory, or cognitive user characteristics that are expected to affect usability;

b) total number of participants tested;

c) user group(s) that were tested and why they were selected;

d) key characteristics and capabilities of each user group;

e) how participants were selected and whether they were representative of the essential characteristics;

f) any differences between the participant sample and the user population;

g) table of participants (row) by characteristics (columns), including relevant demographics, experience, and any special needs.

The characteristics shall be complete enough so that an essentially similar group of participants can be recruited.

Characteristics should be chosen to be relevant to the product's usability; they should allow a reader to determine how similar the participants are to the user population.

EXAMPLE TABLE The table below is an example; the characteristics that are shown are typical but might not necessarily cover every type of testing situation.

	Gender	Age	Linguistic ability	Physical attributes	Task experience	Related products experience
P1						
P2						
Pn						

e.g. for 'Gender', indicate male or female.

e.g. for 'Age', state the chronological age of the participant or indicate membership in an age range (e.g. 25 years to 45 years) or age category (e.g. under 18 years, over 65 years) if the exact age is not known.

e.g. for 'Linguistic ability', state any relevant language skills.

e.g. for 'Physical attributes', describe any physical attributes related to usability.

e.g. for 'Task experience', describe relevant background such as how much experience the user has carrying out this type of task.

e.g. for 'Related product experience', indicate the type and duration of any prior experience with any similar products.

F.4.3 Context of product use in the test

The following information shall be provided:

— any known differences between the evaluated context and the expected context of use.

F.4.3.1 Tasks

The following information shall be provided:

a) task scenarios for testing;

b) why these tasks were selected;

EXAMPLE 1 The most frequent tasks, the most important tasks.

c) how these tasks were identified;

EXAMPLE 2 Observation of consumers using similar products, product marketing specifications.

d) any task data given to the participants;

e) completion criteria established for each task.

F.4.3.2 Test environment

The following information shall be provided:

a) setting and type of space in which the evaluation was conducted;

EXAMPLE 1 Usability lab, meeting room, home office, home family room, public hall.

b) any relevant features or circumstances that could affect the results;

EXAMPLE 2 Video and audio recording equipment, one-way mirrors, or automatic data collection equipment.

F.4.3.3 Technical and technological environment

If the product needs surrounding equipment or technical or technological facilities to function, this shall be specified, e.g. cell phone service provider, TV controlled by a remote control.

NOTE Examples of the technical and technological environment are given in ISO/IEC 25063.

F.5 Experimental Design

F.5.1 General

The following information shall be provided:

— measures for which data were recorded.

F.5.2 Procedure

The following information shall be provided:

a) operational definitions of measures;

b) time limits on tasks;

c) policies and procedures for interaction between tester(s) and subjects.

The following information should be provided:

a) sequence of events from greeting the participants to dismissing them;

b) details of non-disclosure agreements, form completion, warm ups, pre-task training, and debriefing;

c) verification that the participants knew and understood their rights as human subjects;

d) steps followed to execute the test sessions and record data;

e) number and roles of people who interacted with the participants during the test session;

f) whether other individuals were present in the test environment;

g) whether participants were paid or otherwise compensated.

F.5.2.1 Participant general instructions

The following information shall be provided:

a) instructions given to the participants (here or in an Appendix);

b) instructions on how participants were to interact with any other persons present, including how they were to ask for assistance and interact with other participants, if applicable.

The following information should be provided:

a) confirmation that participants knew and understood their rights as human subjects (with the text of the consent form included in an Appendix).

F.5.2.2 Participant task instructions

The following information shall be provided:

— task instruction summary.

F.5.3 Usability measures

The following information shall be provided as specified in Clause 7:

a) measures for effectiveness;

b) measures for efficiency;

c) measures for satisfaction.

The relative importance of each measure may be explained.

F.6 Results

F.6.1 General

The following information shall be provided in sufficient detail to allow replication of the data scoring methods by another organization if the test is repeated:

a) data collection and scoring;

b) data reduction;

EXAMPLE 1 How data were collapsed across tasks or user groups.

c) statistical analysis.

EXAMPLE 2 How any significant differences were calculated.

The following information shall be provided:

— tabular performance results per task or user group including the confidence intervals.

Various graphical formats are effective in describing usability data at a glance. A variety of plots can be used effectively to show comparisons of expert benchmark times for a product versus the mean participant performance time.

F.6.2 Performance results

A table of results may be presented for groups of related tasks where this is more efficient and makes sense.

The following information should be provided:

a) summary table(s) of performance results across all tasks;

b) graphical presentation of performance results.

EXAMPLE TABLE

User #	Task success	Task time (min)	Satisfaction score
1			
2			
N			
Mean			
Confidence interval			
Min.			
Max.			

F.7 Appendices

The following information shall be provided:

a) participant general instructions (if not in the body of the report);

b) participant task instructions.

The following information may be provided:

— release notes explaining or updating the test results.

Annex G
(normative)

Specification of usability requirements

G.1 Introduction

A specification of usability requirements can be used (for example, in procurement or, internally, in development) to specify the requirements for usability and/or accessibility (in terms of effectiveness, efficiency, and satisfaction) of the achievement of user goals (including setup and installation) and the context in which the requirements should be evaluated (the user groups, goals, and environments to be used in testing).

Annex G specifies the information that shall be included in a usability specification.

G.1.1 Overview

Context of use: The description of the context of use shall include the following items:

— users;

— goals;

— physical and social environments.

Usability measures: The usability measures shall include:

— for the most important goals: test scenarios and measures for each scenario;

— at least one target value for each test scenario.

G.2 Context of use

G.2.1 Users

The following information shall be provided:

a) list of expected user groups;

b) key characteristics and capabilities of each user group for which requirements are provided.

Characteristics should be chosen to be relevant to the product's usability; they should be sufficiently detailed to ensure that test participants, who are representative of the actual users, can be selected.

NOTE Relevant factors may include experience in using computers, using similar products, familiarity with the task, frequency of usage, expertise, training, culture, nationality, age, special needs, and the purpose of using the product.

Requirements should be provided for the most important user groups, e.g. the most frequent or critical groups (that could include minor user groups, see C.5).

G.2.2 Goals

The goals for each user group shall be listed, without reference to any specific means of achieving them. The goal should be an intended outcome of value to the user.

NOTE In some cases, it may be possible to identify and list all the usual goals. In other situations, the list may be limited to installation and setup, the most important tasks, representative tasks, or those most frequently encountered.

G.2.3 Physical and social environments

Any aspects of the expected physical and social environments that may influence usability shall be specified:

a) physical environment in which the product will be used, including location and relevant physical conditions, such as temperature or lighting;

b) organizational environment, including group work dynamics, time pressures, supervision, and support;

c) any physical or financial risks, as well as health and safety issues.

G.3 Usability measures

Select which tasks to elaborate with performance and satisfaction goals, for example, those that are most frequent or most critical.

The goals and associated tasks that will form the basis of the requirements shall be specified with test scenarios. For each goal to be elaborated, define one or more specific test scenarios that identify the characteristics of the user and usage environment and the purpose for carrying out the task. There may be separate test scenarios for first use and experienced use.

For each test scenario, appropriate measures shall be identified. These may include the following:

a) Effectiveness: the unassisted completion rate. The completion rate is the percentage of participants who completely and correctly achieve each goal within the predefined maximum time. If goals can be partially achieved (e.g. by incomplete or sub-optimum results), then it may also be useful to set requirements for partial goal achievement, scored on a scale of 0 % to 100 % based on specified criteria related to the value of a partial result (see H.6.1.1).

   EXAMPLE The percentage of customers who can successfully complete a transaction on a website.

   NOTE The requirements for effectiveness are based on correct outcome. It is acceptable that users make errors during the interaction, presuming that they recover from the errors. Keeping track of the errors is particularly helpful in determining what aspects of the product need improvement.

b) Efficiency: the mean time taken to successfully complete each goal and any other measures of resources (see Table 1).

c) Satisfaction: mean score using an established questionnaire.

The relative importance of the measures for each goal should be stated.

Target values are the minimum acceptable value derived from the value for an existing or competitor system value used for the same tasks and goals or based on expert judgment. Additional target values better than the minimum acceptable value may also be specified.

Each target value may be given as

— a definite requirement,

— a provisional requirement subject to further negotiation, or

— an objective for guidance.

NOTE A common way to establish target values is to first measure the effectiveness, efficiency, and satisfaction for the same goals using an existing or competitor system, and then to set the requirements for the new system to be as good as or better than the existing system.

When target values are specified as requirements, they should include the required statistical confidence level (confidence that the result has not occurred by chance) (see 7.5.3).

G.4 Requirements for testing

If requirements for testing are included, the format in Annex H shall be used. Annex H provides the requirements for specifying the method to be used to test whether the usability requirements have been met and the context in which the measurements will be made.

NOTE The performance and satisfaction criteria specified in G.3 are the values required in the intended context of use. The accuracy and precision with which they are measured will be determined by the context of measurement (the number and types of users, the way in which the tasks are simulated, and the simulated working environment).

Annex H
(normative)

Usability test procedure specification

H.1 General

Annex H lists the information to be included when specifying the procedure used to test whether the usability requirements (specified in Annex G) have been met. The results of the test can subsequently be documented using the format in Annex F.

Sufficient information shall be provided to enable the requirements to be tested.

H.2 Users

The following information shall be provided:

a) segmentation of user groups tested, if more than one;

b) total number of participants required;

c) how participants are to be selected so that they have the essential characteristics.

Characteristics should be chosen to be relevant to the product's usability; they should be sufficiently detailed to ensure that the participants are representative of the actual users.

The following information should be provided:

a) any expected differences between the participant sample and the user population;

EXAMPLE The participants' experience with the product will be less than the regular users for whom it is intended.

b) description of any groups with special needs.

H.3 Context of product use in the test

H.3.1 Goals

The following information shall be provided:

a) task scenarios to be tested;

b) criteria for successful completion of each goal;

c) any task data to be given to the participants.

H.4 Test facility

The following information shall be provided:

a) setting and type of space in which the evaluation will be conducted;

EXAMPLE Usability lab, meeting room, home office, home family room, manufacturing floor.

b) how relevant aspects of the intended usage environment will be simulated.

H.4.1 Technical and technological environment

If the product needs surrounding equipment or technical or technological facilities to function, this shall be specified, e.g. cell phone service provider, TV controlled by a remote control.

H.4.2 Ethical issues

Any procedures or regulations to be adhered to for ethical issues (5.5) shall be specified.

H.5 Test procedure

The following information shall be provided:

a) if more than one condition is being tested, the logical design of the evaluation;

b) measures to be taken.

H.5.1 Procedure

The following information shall be provided:

a) operational definitions of measures;

b) operational definitions of any independent variables or control variables;

c) maximum time limits on goals;

d) policies and procedures for interaction between tester(s) and subjects.

H.5.2 Participant general instructions

The following information shall be provided:

a) instructions to be given to the participants (here or in an Appendix);

b) instructions on how participants are to interact with any other persons present, including how they are to ask for assistance or interact with other participants, if applicable.

H.5.3 Participant task instructions

The following information shall be provided:

— task instruction summary.

H.6 Usability measures

The following information shall be provided:

a) measures to be used for effectiveness;

b) measures to be used for efficiency;

c) measures to be used for satisfaction.

Effectiveness, efficiency, and satisfaction shall be measured, even when they are difficult to interpret within the specified context of use. In this case, the requirements shall specify why the supplier does not consider the metrics meaningful.

EXAMPLE Suppose that the context of use for the product includes real-time, open-ended interaction between close associates. In this case, time-on-task may not be meaningfully interpreted as a measure of efficiency because, for many users, time spent on this task is "time well spent".

If it is necessary to provide participants with assists, efficiency and effectiveness metrics shall be provided for both unassisted and assisted conditions and the number and type of assists shall be included as part of the results.

H.6.1 Effectiveness

Effectiveness relates the goals of using the product to the accuracy and completeness with which these goals can be achieved. It does not take account of how the goals were achieved, only the extent to which they were achieved. Efficiency relates the level of effectiveness achieved to the quantity of resources expended.

H.6.1.1 Task completion rate

The completion rate is the percentage of participants who completely and correctly achieve each goal. If goals can be partially achieved (e.g. by incomplete or sub-optimum results), then it may also be useful to set requirements for partial goal achievement, scored on a scale of 0 % to 100 % based on specified criteria related to the value of a partial result. For example, a spell-checking task might involve identifying and correcting 10 spelling errors and the completion rate might be calculated based on the percentage of errors corrected. Another method for calculating completion rate is weighting, e.g. spelling errors in the title page of the document are judged to be twice as important as errors in the main body of text. The rationale for choosing a particular method of partial goal analysis should be stated, if such results are included in the requirements.

The following measure shall be taken:

— percentage of participants who completely and correctly achieve each goal.

H.6.2 Efficiency

H.6.2.1 Measures

Efficiency relates the level of effectiveness achieved to the quantity of resources expended. Efficiency is generally assessed by the mean time taken to achieve the task. Efficiency may also relate to other resources (e.g. total cost of usage). A common measure of efficiency is time on task.

The following measures shall be taken:

— mean time taken to complete each task, together with the range of times across participants.

H.6.3 Satisfaction

The following information shall be provided:

— questionnaire to be used.

Satisfaction describes users' subjective response when using the product. User satisfaction may be an important correlate of motivation to use a product and may affect performance in some cases. Questionnaires to measure satisfaction and associated attitudes are commonly built using Likert and semantic differential scales.

H.7 Statistical significance

The required statistical confidence level (confidence that the result has not occurred by chance) shall be specified (see 7.5.3) for all measures.

H.8 Appendices

The following types of information may be provided:

a) custom questionnaires, if used;

b) participant general instructions (if not in the body of the requirements);

c) participant task instructions.

Annex I
(informative)

Feedback on the standard

I.1 Address for feedback

Comments on the content can be sent to iso_tc159@din.de.

I.2 Feedback invited

Feedback on this part of ISO/TS 20282 is particularly invited on the following issues:

— experiences of usage, case studies, and any published results;

— type of products, area of application, numbers of groups, numbers of users, incorporation of disabilities;

— any problems with the method;

— suggestions for improvements;

— how easy is it to identify the main goals;

— how easy is it to recruit a representative sample of users;

— if two organizations independently test the same product, do they get essentially the same results;

— are the results obtained when testing consumer products reliable and easy to interpret;

— is the method suitable for testing other types of products.

Bibliography

ISO Standards

[1] ISO 8317:2003, *Child-resistant packaging — Requirements and testing procedures for reclosable packages*

[2] ISO/IEC TR 9126-4:2004, *Software engineering — Product quality — Part 4: Quality in use metrics*

[3] ISO 9241-11:1998, *Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability*

[4] ISO 9241-210:2010, *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*

[5] ISO 15535:2006, *General requirements for establishing anthropometric databases*

[6] ISO/TS 18152:2010, *Ergonomics of human-system interaction — Specification for the process assessment of human-system issues*

[7] ISO/IEC 25062:2006, *Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Common Industry Format (CIF) for usability test reports*

[8] ISO/TR 22411:2008, *Ergonomics data and guidelines for the application of ISO/IEC Guide 71 to products and services to address the needs of older persons and persons with disabilities*

[9] ISO/IEC 25010:2011, *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models*

[10] ISO/IEC 25060:2010, *Common industry format for usability – General framework for usability-related information*

[11] ISO/IEC 25062:2006, *Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Common Industry Format (CIF) for usability test reports*

[12] ISO/IEC 25063, *Systems and software engineering — Systems and software product Quality Requirements and Evaluation (SQuaRE) — Common Industry Format (CIF) for usability: Context of use description*

[13] ISO/IEC 25064, *Systems and software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Common Industry Format (CIF) for usability: User needs report*

[14] ISO 26800, *Ergonomics — General approach, principles and concepts*

Other documents

[15] Bangor A., Kortum P., Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. *J. Usability Stud.* 2009, **4** pp. 114–123

[16] Brooke J. SUS: A "quick and dirty" usability scale (1986). In: Jordan, P., Thomas, B., Weerdmeester, B. (eds.) Usability Evaluation in Industry, pp. 189–194. Taylor and Francis.

[17] Hassenzahl M.AttrakDiff(tm). Available from http://www.attrakdiff.de

[18] Jäger R. Konstruktion einer Ratingskala mit Smilies. *Diagnostica*. 2004, **50** pp. 31–38

[19] Jian J.-Y., Bisantz A.M., Drury C.G. Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergon.* 2000, **4** pp. 53–71

[20] Kunin Th. The construction of a new type of attitude measure. *Person. Psychol.* 1955, **8** pp. 65–77

[21] Nielsen J. *Usability engineering*. Morgan Kauffman, 1993

[22] Norman K., Alm H., Wigaeus Tornqvist E., Toomingas A. Reliability of a questionnaire and an ergonomic checklist for assessing working conditions and health at call centres. *Int. J. Occup. Saf. Ergon.* 2006, **12** pp. 53–68

[23] Power C., Freire A., Petrie H., Swallow D. 2012). Guidelines are only half of the story: accessibility problems encountered by blind users on the web. In Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, 433-442.

[24] Sauro J. (2011) Measuring Usability With The System Usability Scale (SUS). Available from http://www.measuringusability.com/sus.php

[25] Sauro J. (2010) Confidence Interval Calculator for a Completion Rate. Available from http://www.measuringusability.com/wald.htm

[26] Sauro J., & Dumas J.S. (2009) Comparison of Three One-Question, Post-Task Usability Questionnaires. Proceedings of CHI 2009.

[27] Sauro J., & Iklund K. (2004) How long should a task take? Identifying Spec Limits for Task Times in Usability Tests. Available from http://measuringusability.com/time_specs.htm

[28] Sauro J., & Lewis J. (2005) Estimating completion rates from small samples using binomial confidence intervals: comparisons and recommendations. Proceedings of the Human Factors and Ergonomics Society 49th annual meeting.

[29] Sauro J., & Lewis J.R. (2010) Average Task Times in Usability Tests: What to Report? in Proceedings of the Conference in Human Factors in Computing Systems (CHI 2010) Atlanta, GA.

[30] Sauro J., & Lewis J.R. *Quantifying the User Experience: Practical Statistics for User Research*. Morgan Kaufmann, 2012

[31] Tullis T. (2008) Calculating a confidence interval for task completion using the Adjusted Wald Method. Available from http://measuringux.com/AdjustedWald.htm

[32] Zijlstra F.R.H., & van Doorn L. *The construction of a scale to measure subjective effort. Technical Report*. Delft University of Technology, Department of Philosophy and Social Sciences, 1985

ICS 13.180

Price based on 50 pages