

I n t e r n a t i o n a l T e l e c o m m u n i c a t i o n U n i o n

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1501

(02/2014)

**SERIES P: TERMINALS AND SUBJECTIVE AND
OBJECTIVE ASSESSMENT METHODS**

Methods for objective and subjective assessment of
quality of services other than voice services

**Subjective testing methodology for web
browsing**

Recommendation ITU-T P.1501

ITU-T



ITU-T P-SERIES RECOMMENDATIONS
TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Voice terminal characteristics	Series	P.30
		P.300
Reference systems	Series	P.40
Objective measuring apparatus	Series	P.50
		P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of speech quality	Series	P.80
		P.800
Audiovisual quality in multimedia services	Series	P.900
Transmission performance and QoS aspects of IP end-points	Series	P.1000
Communications involving vehicles	Series	P.1100
Models and tools for quality assessment of streamed media	Series	P.1200
Telemeeting assessment	Series	P.1300
Statistical analysis, evaluation and reporting guidelines of quality measurements	Series	P.1400
Methods for objective and subjective assessment of quality of services other than voice services	Series	P.1500

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T P.1501

Subjective testing methodology for web browsing

Summary

Recommendation ITU-T P.1501 describes subjective testing methods for assessing the user perceived quality for web browsing in browser-based applications of different device classes. It gives guidance on the selection of a test environment, equipment and content, and how test procedures should be properly planned in order to facilitate realistic web browsing behaviour by the subjects. The methods described can be used to identify the impact of several different factors that influence the user perceived quality of web browsing, e.g., identifying the influence of the browsing device on user perceived quality for identical technical access network settings.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.1501	2014-02-13	12	11.1002/1000/12126

Keywords

Absolute Category Rating, Mean Opinion Score, page view cycle, web browsing, web-QoE.

* To access the Recommendation, type the URL <http://handle.itu.int/> in the address field of your web browser, followed by the Recommendation's unique ID. For example, <http://handle.itu.int/11.1002/1000/11830-en>.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2014

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

	Page
1 Scope.....	1
2 References.....	1
3 Definitions	1
3.1 Terms defined elsewhere	1
3.2 Terms defined in this Recommendation.....	2
4 Abbreviations and acronyms	2
5 Conventions	2
6 Web browsing test procedure	2
6.1 Purpose	2
6.2 Test facilities	3
6.3 End-user devices.....	4
6.4 Test content	4
6.5 Test procedure	5
6.6 Test conditions.....	6
6.7 Subjects.....	6
6.8 Tasks.....	6
6.9 Questions	7
6.10 Data analysis and report	7
Annex A – Result reporting: mandatory structure and details.....	9
A.1 Test set-up	9
A.2 Result presentation	10
Appendix I – Example of instructions for the web browsing test.....	11
Appendix II – Example of task instructions for the free-browsing test	12
Appendix III – Example of instructions for the task-dependent browsing test	14
Appendix IV – Example of questions for the web browsing test	15
IV.1 Assessment of media quality	15
IV.2 Assessment of temporal aspects	15
Appendix V Example test plan	16
V.1 Test facilities	16
V.2 Test content	16
V.3 Test conditions and test design.....	16
V.4 Result presentation	17
Bibliography.....	18

Recommendation ITU-T P.1501

Subjective testing methodology for web browsing

1 Scope

This Recommendation describes the methods and procedures used for subjective testing of user perceived quality of web browsing. The method generally applies to degradations and characteristics that can be introduced at the network level (e.g., round-trip time, downlink and uplink bandwidth, packet losses), as well at the application level (e.g., page load-times). Combinations of two or more of these factors also have to be considered. Impacting factors due to webpage design are not in the scope of this Recommendation.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

- [ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.
- [ITU-T P.800.1] Recommendation ITU-T P.800.1 (2006), *Mean Opinion Scores (MOS) terminology*.
- [ITU-T P.910] Recommendation ITU-T P.910 (2008), *Subjective video quality assessment methods for multimedia applications*.
- [ITU-T Y.1540] Recommendation ITU-T Y.1540 (2011), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.
- [ITU-T Y.1541] Recommendation ITU-T Y.1541 (2011), *Network performance objectives for IP-based services*.
- [ITU-R BT.500] Recommendation ITU-R BT.500 (2012), *Methodology for the subjective assessment of the quality of television pictures*.
- [ITU-T HSTP] ITU- T Handbook (2011), *Practical procedures for subjective testing*, ITU, Geneva.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 condition: Time period in which a defined quality of service (QoS) setting is applied for a certain webpage. A condition is concluded by retrospective user ratings on one or multiple dimensions of quality.

3.2.2 element: Visual content of a webpage, which is displayed to the user on the rendered webpage, e.g., text, pictures, widgets, videos.

3.2.3 object: HTTP object, which is used for processing and rendering the webpage and is referenced by the page mark-up or script. These objects are not necessarily visible on the fully rendered page (e.g., elements).

3.2.4 task: A set of instructions given to the user on how to reach certain information within a test condition.

3.2.5 web browsing session: An interactive information exchange between a user and one or more websites over a limited period of time, mediated via a web browsing application. The starting point of a session is the first page request initiated by the user, which is followed by a number of request-response interactions between the user and the webhost(s), resulting in a series of page views. See [b-Egger-2]. A web session is typically terminated when the user exits the browsing application or stops the browsing activity for a certain period of time.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

DLBW Downlink Bandwidth

MOS Mean Opinion Score

RTT Round Trip Time

ULBW Uplink Bandwidth

WBS Web Browsing Session

5 Conventions

The methodology discussed in this Recommendation is tailored for web browsing only. Browser-based applications, such as online office applications, webmail, etc., might be evaluated by the same means, but they have to consider different scenarios depending on the defined tasks during subjective evaluation and the technical implementation of the services.

The focus of this Recommendation is the evaluation method of the complete web browsing session (WBS) to achieve and maintain a flow experience of web browsing for better external validation of the test results. However, this methodology's test set-up with respect to test facilities, test conditions, subject selections and test devices is also applicable to the subjective evaluation of single web page loading times.

6 Web browsing test procedure

6.1 Purpose

Subjective quality tests for web browsing aim to assess the impact of transmission impairments on the user-perceived quality of a web browsing session (WBS) under realistic usage conditions.

The main characteristics of a subjective web browsing test are:

- To simulate realistic web browsing where users are browsing and interacting with webpages in order to acquire certain information; the procedure they go through within this methodology ensures that users get into a browsing mode rather than a pure page loading mode.
- Subjects are exposed to a certain quality of service (QoS) level over a period of time rather than for one event, in order to grasp several request-response cycles for the subjective evaluation.
- The use of certain tasks to stimulate the interaction between the webpage and the subject for each test condition.
- The webpage must be interactive and has to provide sufficient content such that the subject can browse through it over several conditions, without getting bored.
- This test methodology can be adapted to field testing; however, it has to be ensured that the manipulated parameters (e.g., delay, packet loss, downlink bandwidth) can be set to the desired values and that these settings can be verified by a posteriori analysis (e.g., traffic traces).

The results obtained with this methodology can be used by telecommunication providers, vendors and customers to get information on the influence of the tested technical parameters on end-user perceived quality.

6.2 Test facilities

Typically, test facilities for subjective web browsing evaluations include an end-user device, a network emulator, a router and some optional components as depicted in Figure 6-1 below. Network parameters are set through a network emulator (EMU) which is placed between the user device (UD) and the router (RT), which ensures the eventual routing between local content servers and the internet.

When using live webpages within the evaluation one has to ensure the measurement of the page health, in order to identify potential influences due to congestion or server malfunctions. This is done through a page health monitor (PHM) connected to the Internet via the router. See Figure 6-1. In the case of multiple subjects participating in parallel, it must be ensured that each subject receives the obliged parameters, hence shared channel settings have to be avoided.

If webpages are used, that include video content that is also affected by the test parameters, additional requirements for the lab facilities as specified in [ITU-T P.910] apply.

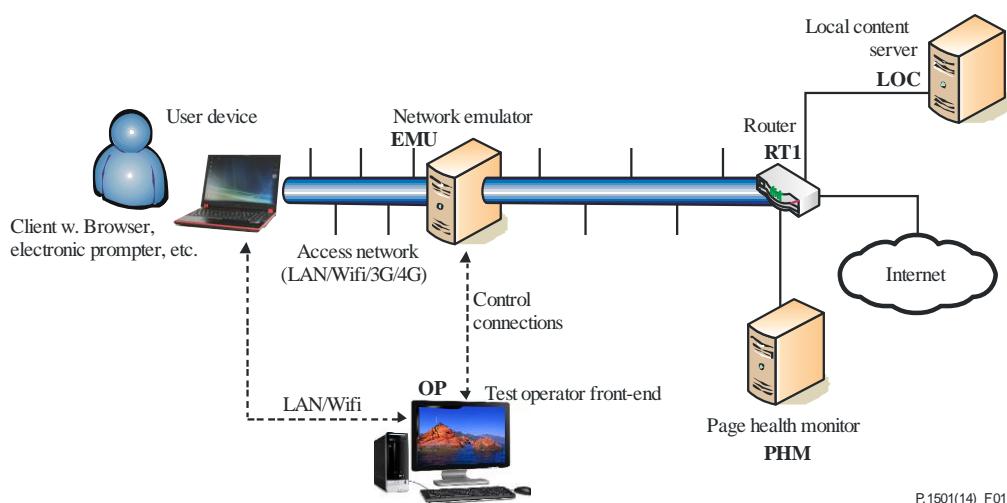


Figure 6-1 – Example of test facilities

6.3 End-user devices

The end-user device has a major impact on the overall performance as perceived by the user. Therefore, detailed reporting about the device characteristics and setting are essential. If the end-user device or only any of its characteristics are considered to be a variable parameter of the quality evaluation, then all applied variations must be reported.

6.3.1 Hardware

Due to the great dynamics in web technology developments, it is desirable that modern hardware is used for the tests in order to ensure that unintended device performance does not dominate the chosen test variables.

At the time of drafting this Recommendation, hardware can be roughly classified in four categories: 1) PC/laptop, 2) tablets, 3) smartphones and 4) other devices that are able to connect to the web. Typically, these four categories feature different performance issues. Therefore, one should only use devices from different categories when they are part of the condition variables under test.

Displays should be calibrated to the D65 standard illuminant at the beginning of the test if possible (Note, this might not apply to device categories 2, 3 and 4; subjects must not be allowed to modify the display's luminance level).

For each of the devices used within a test, the technical details have to be included in the reporting as detailed in Annex A.

6.3.2 Software

Regarding the operating system of the devices, the latest updated versions of their major operating systems should be used. The software related details that have to be included in the report are listed in Annex A

6.3.2.1 Browser

For the device categories 2, 3 and 4, it is recommended to use their default browser as these are optimized in several ways for these devices. In case of category 1, the latest version of a major browser should be used. However, one might apply older legacy browsers, if needed to run special browser extensions. Regarding these browser extensions, as few extensions as possible should be used to ensure the performance of the browser.

6.3.2.1.1 Browser cache

At the start of each condition the browser cache should be cleared so that the WBS starts at a defined status. However, throughout the WBS, the browser cache has to be enabled to ensure realistic loading behaviour of webpages.

6.4 Test content

Typical webpages should be chosen when selecting content. One has to ensure that the webpage is well programmed and its server infrastructure is fast and reliable across peak network times. It is recommended that server performance is monitored over a one week period before a webpage is chosen as test content.

Content types and examples follow: news page (e.g., www.news.de), photo album (locally hosted), web shopping (e.g., www.shop-on-line.de). Content examples are chosen such that they reflect typical webpages frequently browsed by typical Internet users. In the section below, categorization of webpages is described.

6.4.1 Webpage categorization

Categorization of webpages is best performed according to technical page complexity. The complexity of a webpage can be best captured by the number and size of objects fetched to load the

webpage and also the different MIME types (e.g., image, javascript, CSS, text) across which these objects are spread (see [b-Butkiewicz]).

Since technical page complexity leads to a multi-dimensional continuum rather than to discrete clusters, webpages cannot be easily categorized. Therefore, to ensure sufficient information about the used content, the following statistics for each visited webpage, and its subpages, have to be reported (see clause A.1.3):

- average page size in kBytes and standard deviation of the page size;
- average number and size of elements per page and per MIME type, and standard deviation of number of elements and size per page and per MIME type;
- average number of objects and size (in the HTML source code of the page) per page and per MIME type, standard deviation of number of objects and size per page and per MIME type.

6.5 Test procedure

The flow characteristics of web browsing demand condition lengths with several technical request-response patterns (see [b-Egger-2]). Therefore, a condition length between 60 and 120 seconds is recommended, followed by the retrospective ratings on the dimensions of overall quality of the connection. Including the ratings, one condition is of about 80 seconds and 140 seconds in length respectively. Such a length is a reasonable trade-off between achieved user flow and practical considerations such as the number of possible conditions and the maximum overall test duration. A maximum active test duration of two 45-minute blocks separated by a 10-minute break is recommended (see [b-Schatz]). In addition, an instruction block at the beginning and a debriefing block at the end are also recommended. The overall duration should not exceed 120 minutes (see [b-Schatz]). An example of a test schedule with 24 test conditions is shown in Table 6-1.

Compromises have to be made between the test duration and the choice of conditions. If more conditions are to be tested, different experiments may require a between-subject or balanced incomplete block design. The order of the conditions should be pseudo randomized, e.g., by means of Latin squares or other experimental designs as described in section 3.3 of [ITU-T HSTP].

Each test subject should be seated in a realistic environment according to the chosen scenario. The beginning and end of each condition must be clearly indicated to the test subject to ensure ratings corresponding to the experimental parameters set. However, care has to be taken to make these notifications clear, but also not too intriguing in order to keep the subject in the web browsing flow.

After arrival of the subjects, instruction has to be provided to explain the nature of the test and tasks. This should be followed by a training session in which the participant goes through all the different tasks given, to ensure correct understanding and execution of the tasks in which he or she practices the use of the rating scales to rate the quality retrospectively. The training session should also ensure that parameter values from the lower end and the higher end of the parameter range are used in order to acquaint the subjects with the parameter range to be used throughout the test. In addition, it has to be ensured that the parameter ranges are chosen such that the whole range of the scale is used.

Table 6-1 – Test schedule for 24 conditions based on 60-second condition length

	Instruction	Training session	Session 1	Break	Session 2
Number of test conditions		3	12		12
Time	15 minutes	4 minutes	16 minutes	10 minutes	16 minutes

6.6 Test conditions

Typically for web technologies, services are delivered via IP networks. These networks are currently available in fixed and mobile scenarios with different network characteristics and are accessible using various devices. For the design of the test this has to be considered in order to achieve a realistic scenario for the test. A more detailed discussion regarding IP parameters can be found in [ITU-T Y.1540] and [ITU-T Y.1541]

Examples of test parameters are:

On a network level:

- downlink bandwidth (DLBW)/uplink bandwidth (ULBW)
- round trip time (RTT)
- temporal fluctuations of DLBW, ULBW, RTT
- etc.

On an application level:

- page load time

On a device level:

- devices (PC, laptop, tablet, smartphone)

6.7 Subjects

Special care should be taken when selecting subjects for tests with tablet and smartphone devices. These devices are not as widespread as PCs and some potential subjects will be more experienced than others. It is recognized that the level of experience with specific equipment or technology is a continuum ranging from those who are completely unfamiliar with the technical behaviour of the equipment under test (i.e., "non-experts") to those who are thoroughly competent in the operation and maintenance of this equipment (i.e., "experts"). Another criterion for subject selection to be considered in web browsing is the subject's experience and regular usage of the test content. It is noted that good knowledge about the structure and usual behaviour of the test webpage can have a severe impact on the sensitivity due to test parameter changes, measured for test subjects.

In order to ensure that the influence of the above-mentioned experience with the devices and webpages under test are minimized, it is recommended that the selected subjects are experienced with the devices under test and the type of webpages used. This means that the potential subjects should be using the device type or accessing the type of webpage at minimum on a weekly basis.

Unless gender, age and other socio-economic characteristics are design factors of the test, then a formal web browsing test should be populated (on a best-effort basis) with a random mix of subjects.

The age and gender as well as socio-economic characteristics of all types of subjects should be recorded for all test types.

6.8 Tasks

In general, two different modes for the task during a WBS can be defined, free-browsing mode and task-dependent mode. The free-browsing mode does not require specific goals to be achieved during the session. Subjects are asked to explore the test content on their own by requesting several webpages. In task-dependent mode, specific tasks are defined that the subject shall perform on the webpages. The end of such sessions can be determined either by a specific number of webpages requested and loaded consecutively, by achieving the task goal or by a given condition time.

In general, the task shall be intrinsically motivating and easily solvable without training. Additionally, the task accomplishment must not be limited by existing/known usability problems to avoid unintended impact of task design on test parameters. To avoid boredom, tasks should include a logical

structure and relationship to the purpose of the website (e.g., look for offers on shopping page or look for information on the latest headline on a news page). Examples of suitable tasks for different website classes are included in Appendix II for the free-browsing mode and in Appendix III for the task-dependent mode.

6.9 Questions

Single or multiple dimensions of perceived quality are evaluated after each test condition. Examples of absolute category scale questions and labels as well as a binary response scale are given below. It is noted that the actual questions chosen for the test are dependent on the research questions to be answered and might exceed these questions. Further examples of questions for other dimensions of interest are given in Appendix IV. However, it has to be considered that the cognitive load on the subjects and therefore the number of questions asked should be kept to a minimum in order to reduce any possible confusion and subject fatigue.

6.9.1 General assessment of the perceived overall quality of a test item

6.9.1.1 "Please rate the overall quality of the internet connection using the following scale"

The scale descriptors are:

- Excellent
- Good
- Fair
- Poor
- Bad

6.9.2 Identification of quality thresholds with respect to acceptance of the condition

Please note that a context must be defined for the threshold (e.g., accepted quality for your internet connection at home; accepted quality for the price paid for the service). The context reference is indicated by the underlined words in the example question.

6.9.2.1 "Would you accept this connection quality in normal life at home?"

- Yes
- No

The previous examples should be supplemented by the experimenter to address the needs of the specific experiment. When using multiple scales for assessing the multi-dimensional aspect of quality, care should be taken to ensure that the previous responses are not available to the subjects.

6.10 Data analysis and report

6.10.1 Analysis methods

Regarding the analysis methods, the recommendations from [ITU-T P.800], [ITU-T P.805] and section 6 of [ITU-T HSTP] apply.

Considering the small sample sizes achieved, one can, in addition to the analysis methods mentioned in the above references, also use non-parametric statistical hypothesis tests such as Mann-Whitney U, Wilcoxon signed rank tests and others.

6.10.2 Reporting on results and test design

The report must contain the questions and the category scales used. Due to the possible influence from the set-up itself, it is also mandatory to report the following set-up facts in detail: webpages tested, browser name and version of the browser used, plugins used, and date and time of the test.

Summary results should include, at a minimum, mean ratings and standard deviations for all tested conditions and for all questions. If the experimenter expects a gender pairing effect, this special case should also be considered in the design of the experiment and particularly in the pairing of subjects. This effect should be analysed and reported.

A list of mandatory structure and reporting details is included in Annex A.

Annex A

Result reporting: mandatory structure and details

(This annex forms an integral part of this Recommendation.)

A.1 Test set-up

A.1.1 Test participants

A description of the participant sample used has to include at least the following details (a more detailed description is of course possible):

- number
- mean age
- median age
- gender

A.1.2 Test parameters

In this section, all independent variables, which were varied throughout the duration of the test, have to be listed with their respective parameter ranges.

A.1.3 Webpages used

For the webpages used the following details have to be reported:

- URL¹
- start-date and end-date of the test¹
- webpage class (e.g., news, photo album, geo location, web shopping)
- web technology used (e.g., HTML, Ajax, Flash, Java)

A.1.4 Test facilities

The test facilities such as the environment where the tests were conducted and the technical set-up used have to be reported in this section such that a replication of the test facilities can be achieved by a third party.

A.1.5 Test devices

For the devices used throughout the test the following details have to be reported:

Hardware:

- brand
- model number
- CPU (type, speed)
- RAM (MB/GB)
- screen size (diagonal)

Screen resolution (i.e., pixel horizontal × pixel vertical) and additionally, in case the browser window is not maximized, its resolution has to be reported.

¹ This has to be reported only in the case where live webpages were used; when local webpages were used as test content these details are not mandatory.

Software:

- operating system and build number
- browser and version number
- enabled browser extensions and version numbers
- speed optimization used (e.g., http pipelining)

A.1.6 Questions used

All question used for the evaluation of the user perceived quality on several dimensions have to be reported in their exact wording in this section. The scale types used, absolute category rating (ACR) (ACR-5, ACR-9, ACR-11) (see [ITU-T P.910]) have to be included in the report as well.

A.2 Result presentation

A.2.1 Acceptance results

For each condition tested, the following details have to be reported:

- number of positive ratings
- number of negative ratings
- number of no responses

A.2.2 Mean opinion score results

For each condition tested, the following mean opinion score (MOS) details have to be reported:

- mean
- number of responses
- standard deviation

For graphical result presentation confidence-intervals have to be shown and their properties (e.g., percentiles) have to be reported.

Appendix I

Example of instructions for the web browsing test

(This appendix does not form an integral part of this Recommendation.)

INSTRUCTIONS TO SUBJECTS

In this experiment, we are evaluating the connection quality of internet connections. Therefore, you are asked to browse through various webpages, solve different tasks and rate the perceived connection quality. Please concentrate on the quality of the internet connection and do not get biased by the content of the webpages.

After completing each condition, you will be asked to give your opinion on the connection quality by answering the questions displayed on the laptop in front of you. Please answer these questions intuitively. The scale is ranked from "Bad" to "Excellent", where "Bad" is the worst and "Excellent" is the best connection quality that you can imagine. After clicking the "save" button, your judgement will be stored. After filling out the form, please continue with the task, until you receive further instructions.

Below you can see an example reply prompt to all questions you will be asked.

Please rate the overall quality of your connection using the following scale

Bad	Poor	Fair	Good	Excellent
-----	------	------	------	-----------

"Would you use/accept this connection quality in normal life at home?"

No	Yes
----	-----

You will have a break after approximately 45 minutes. In total, the test will last approximately 120 minutes.

Appendix II

Example of task instructions for the free-browsing test

(This appendix does not form an integral part of this Recommendation.)

Photos

Tasks

- 1) Please choose one of the photo albums on the start page.
- 2) Select the first picture preview and wait until it is completely displayed.
- 3) Browse through the album by clicking on the "next" button or click directly on the picture. Please view each picture for at least three seconds.
- 4) For each photo task, note the number of the five pictures you like most and the five you like least.

News page

Task

Try to get an overview of the current news in the economy, politics, sports and culture sections. Also, click on the articles to get an **impression** of the topics covered (but please do not read the articles completely). Please keep in mind that you are asked to rate the quality of the internet connection; thus, you should be able to form an opinion about it. Please continue reading articles in **one result**, until you get the directive to select the next result. Please stay on the **NEWS** section of the page!

Comment: ideally you should browse through two to three articles per minute

Please **do not surf the video or the picture section of the page**; we are just testing news-articles (i.e., text and pictures)!

Navigation tasks

Imagine you are planning a weekend trip to Vienna, with an overnight stay at a hotel and a little bit of sightseeing.

- 1) Search for "**Vienna**" (by using the search field).
- 2) Look for a hotel on the **Kärntnering** in Vienna.
- 3) Hint: Insert "hotel, Kärntnerring" into the search field.
- 4) Browse through the hotels found (by zooming and panning).
- 5) Look for the hotel which is nearest to "**Minoritenkirche**".
- 6) Find the "**Naturhistorische Museum**", the "**Bestattungsmuseum**" and the "**Arsenal**" in Vienna.
- 7) Get the directions between the named museums. The start and endpoint is your hotel.
- 8) To get the directions, click on the "get directions" button on the right side of your screen. Choose the starting point and add as many destinations as you'll need – right below the input field, is an "add destination" link. Finally, click on "Get directions".
- 9) Zoom in so that you will be able to see the street names (by using the mouse wheel or the "+" at the left top of the map).
- 10) Explore the route (by panning).
- 11) Change the destinations. Start at your hotel and get the direction to "**Donau City Str. 1**".

Web shopping

Please search for literature of choice. Suggestions:

- 1) Look for holiday reading material.
- 2) Search for an interesting book about Barack Obama.
- 3) Find a serious book regarding the reasons for the financial crisis.
- 4) Check the current Amazon charts for literature.
- 5) Search for a tourist guide for your next holiday.
- 6) Find the new book by Ken Follett.
- 7) Look for a self-help book on the subject of "Burn Out".
- 8) Look for a cookbook with Spring recipes.

Warning: Please **DO NOT** watch any videos!

Appendix III

Example of instructions for the task-dependent browsing test

(This appendix does not form an integral part of this Recommendation.)

This appendix contains the instructions given to the test subjects before a test in case of task-dependent browsing mode.

Test instructions

- 1) After you have started the test you will see a news page. Your task is to identify the soccer team of the week. On the first page it is your task to navigate to the sports section in order to identify the soccer team of the week. To proceed with this task please click on "Sports".
- 2) After clicking you will see the sports page. As your task is to identify the soccer team of the week you have to navigate into the soccer section in the right column by clicking "Soccer".
- 3) As soon as you see the soccer section you will find a link to the "team of the week" in the right column. Please click this link in order to proceed with your task.
- 4) After clicking on the "team of the week" link, you will arrive at the team of the week page. You can see the current standings of the soccer league, soccer news and also the information on the "team of the week" at the lower right corner of the screen. Please click this link in order to finish your task.
- 5) After clicking this link a pop-up window appears in which you should select the "team of the week" from the soccer standings and you are asked to rate the quality of the web session you have gone through.
- 6) If you have further questions on the test procedure please feel free to ask the test supervisor.

Appendix IV

Example of questions for the web browsing test

(This appendix does not form an integral part of this Recommendation.)

IV.1 Assessment of media quality

Examples, other than those listed here, can be found in [ITU-T P.910] and [ITU-R BT.500]

IV.1.1 "Please rate the image quality using the following scale"

The scale descriptors are:

- Excellent
- Good
- Fair
- Poor
- Bad

IV.1.2 "Please rate the video quality using the following scale"

The scale descriptors are:

- Excellent
- Good
- Fair
- Poor
- Bad

IV.2 Assessment of temporal aspects

IV.2.1 "Please rate the speed of the page rendering using the following scale"

The scale descriptors are:

- Excellent
- Good
- Fair
- Poor
- Bad

IV.2.2 "Please rate the variability of the page rendering speed using the following scale (Excellent = very low variability, Bad = high variability)"

The scale descriptors are:

- Excellent
- Good
- Fair
- Poor
- Bad

Appendix V

Example test plan

(This appendix does not form an integral part of this Recommendation.)

V.1 Test facilities

For the test facilities (e.g., client devices, network emulator) and the respective set-up the guidelines from clause 6.1 should be followed.

V.2 Test content

The news page and the photo gallery page from [b-Egger-1] should be used as test content. Both webpages can be locally hosted on a dedicated webserver in order to rule out any interference due to network or Internet problems. Respective task descriptions can be found in Appendix II.

V.3 Test conditions and test design

The technical test conditions are described in Table V.1. Each of the test conditions should be presented twice to the user such that 24 total test conditions are achieved and each condition is rated twice. The order of the conditions has to be randomized so the order effects can be ruled out. Please note that the DLBW is just an example and should be chosen according to the user group and their expectations and experiences with DLBWs for Internet connections.

Table V.1 – Technical conditions for both content types

Downlink BW [kbit/s]	64	128	256	512	1024	2048	Conditions
News page	x	x	x	x	x	x	6
Photo page	x	x	x	x	x	x	6
RTT (fixed)	20 ms					Total	12

For the test design, each condition should last for 60 seconds and then the user should be prompted with the questionnaire as described in Table V.3 (the average time for the questionnaire is 30 seconds, resulting in a total of 90 seconds per condition, including rating). The time schedule for the test is described in Table V.2. For the warm-up condition, one condition should be the slowest DLBW (64 kbit/s) and one condition should be the fastest DLBW (2048 kbit/s). The third warm-up condition can be randomly chosen from the remaining DLBW conditions (128, 256, 512, 1024 kbit/s). Also the content of the warm-up conditions can be randomly chosen.

Table V.2 – Test design and timings

Task	Duration (s)	Conditions
Instructions	300	
Warm-up	270	3
Webpage 1/2	1080	12
Break	600	
Webpage 2/2	1080	12
Total	3330	27
		56 min

The questions asked after each condition are described in Table V.3.

Table V.3 – Questions for the test participants

Please rate the overall quality of your connection using the following scale:

Bad	Poor	Fair	Good	Excellent
<input type="radio"/>				

Would you accept this connection quality in normal life at home?

Yes	No
<input type="radio"/>	<input type="radio"/>

V.4 Result presentation

The results should be presented as described in Annex A.

Bibliography

- [b-Schatz] R. Schatz, S. Egger and K. Masuch (2012), *The Impact of User Fatigue and Test Duration on the Reliability of Subjective Quality Ratings*. Journal of the Audio Engineering Society, Vol. 60, March; pp. 63-73.
- [b-Butkiewicz] Butkiewicz, Michael, Harsha V. Madhyastha and Vyas Sekar (2011), *Understanding Website Complexity: Measurements, Metrics, and Implications*. Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, 313–328. IMC'11. New York, NY, USA: ACM, 2011. doi:10.1145/2068816.2068846.
- [b-Egger-1] Sebastian Egger, Raimund Schatz (2012), *Interactive Content for Subjective Studies on Web Browsing QoE: A Kepler Derivative*. ETSI STQ Workshop on Selected Items on Telecommunication Quality Matters, 27-28 Nov, Vienna, Austria.
- [b-Egger-2] S. Egger, T. Hossfeld, R. Schatz, M. Fiedler (2012), *Waiting times in quality of experience for web based services*, Fourth International Workshop on Quality of Multimedia Experience (QoMEX), July; pp. 86-96.

SERIES OF ITU-T RECOMMENDATIONS

- Series A Organization of the work of ITU-T
- Series D General tariff principles
- Series E Overall network operation, telephone service, service operation and human factors
- Series F Non-telephone telecommunication services
- Series G Transmission systems and media, digital systems and networks
- Series H Audiovisual and multimedia systems
- Series I Integrated services digital network
- Series J Cable networks and transmission of television, sound programme and other multimedia signals
- Series K Protection against interference
- Series L Construction, installation and protection of cables and other elements of outside plant
- Series M Telecommunication management, including TMN and network maintenance
- Series N Maintenance: international sound programme and television transmission circuits
- Series O Specifications of measuring equipment
- Series P Terminals and subjective and objective assessment methods**
- Series Q Switching and signalling
- Series R Telegraph transmission
- Series S Telegraph services terminal equipment
- Series T Terminals for telematic services
- Series U Telegraph switching
- Series V Data communication over the telephone network
- Series X Data networks, open system communications and security
- Series Y Global information infrastructure, Internet protocol aspects and next-generation networks
- Series Z Languages and general software aspects for telecommunication systems