TECHNICAL SPECIFICATION

ISO/TS 14198

Second edition
2019-04

Road vehicles — Ergonomic aspects of transport information and control systems — Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems

Véhicules routiers — Aspects ergonomiques des systèmes d'information et de contrôle du transport — Tâches de calibration pour méthodes qui évaluent la distraction du conducteur due à l'utilisation des systèmes embarqués

ISO/TS 14198:2019(E)

COPYRIGHT PROTECTED DOCUMENT

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 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso .org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 22, *Road vehicles*, Subcommittee SC 39, *Ergonomics*.

This second edition cancels and replaces the first edition (ISO/TS 14198:2012), which has been technically revised. The main changes compared to the previous edition are as follows:

— in addition to the Lane Change Test (LCT), the Detection Response Task (DRT) is added as a primary task;

— in addition to the Critical Tracking Task (CTT) and Surrogate Reference Task (SURT), the n-back task is added as calibration task.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

The number of standardized methods to assess driver attentional demand due to the use of in-vehicle information and communication devices is continuing to increase. In applying these methodologies, it is important to understand and document variability in participants' performance of standard calibration tasks and procedures across laboratories and/or time.

A suitable calibration task should have the following attributes:

— It should be robust against the variations in cultural background of participants.

— Properly applied, the task should give repeatable quantitative results. It should be sensitive to inappropriate variations in participants, equipment, location, experimenter and instruction.

— It should use durable and readily available equipment for conducting the task.

— It should apply to the driver population and be usable in a driving-like context.

A standardized calibration task can be used to produce a range of statistically stable, repeatable and comparable secondary task demands for a participant in an experimental setting. This setting can be used to assess the effect on driving performance of the attentional demand due to driver interaction with an information, entertainment, and control or communication system while a vehicle is in motion.

Different calibration tasks are specified in this document to cover calibration manual, visual, and cognitive aspects of various secondary task characteristics.

Road vehicles — Ergonomic aspects of transport information and control systems — Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems

1 Scope

This document provides procedures that can be used as a secondary task in a dual-task setting to determine whether that evaluation setting is standardized and valid for purposes of assessing driver attentional demand due to the use of an in-vehicle system. This document does not define calibration procedures for other evaluation activities that a laboratory might undertake.

This document provides guidance on selecting a calibration task given a specific primary task. The primary tasks of interest include those that would be used in the evaluation of attentional demand. Such primary tasks are defined in other documents.

The description of a calibration task includes its application, experimental setup, data collection, and procedures for analysis of results.

The purpose of this document is not to define a reference criterion as to whether a given secondary task is suitable for use while driving. Although specific settings of parameters of a calibration task might be used to realize such a predefined pass/fail criterion, this document does not provide such a criterion for a given level of attentional demand.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 26022:2010, *Road vehicles — Ergonomic aspects of transport information and control systems — Simulated lane change test to assess in-vehicle secondary task demand*

ISO 17488, *Road vehicles — Transport information and control systems — detection-response-task (DRT) for assessing attentional effects of cognitive load in driving*

ISO/IEC 7498-1, *Information technology — Open Systems Interconnection — Basic Reference Model: The Basic Model*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 7498-1 and the following apply.

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

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

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

3.1
calibration task
type of reference task used for the purpose of comparing different tests or test results between sites, or over time at a given site

3.2
criterion
threshold or value of a variable to be met

3.3
demand
total visual, auditory, cognitive, or physical resources required of the driver to accomplish the primary task and interact with a Transport Information and Control System (TICS) in a dual task setting

3.4
dual task
two tasks concurrently performed, typically the primary task plus the secondary task

3.5
environment
physical surroundings in which data are captured and collected, consequently, the level of control over the independent variables in a study

EXAMPLE   Laboratory, simulator, test track, real road.

3.6
evaluation
procedure in which the effect of a Transport Information and Control System (TICS) or another device is assessed

Note 1 to entry: It may be undertaken retrospectively after the TICS has been in use for a considerable time as a product.

Note 2 to entry: The results of the evaluation will depend on the HMI, but also on the equipment reliability and subsequent behavioural changes which may affect driving performance.

3.7
method
high-level approach to an assessment, based on theory and principles, which implies an underlying rationale in the choice of assessment techniques

EXAMPLE   Behaviour analysis, workload assessment, and analysis of psycho-physiological responses.

3.8
metric
quantitative measure of driver behaviour independent of the tool used to measure it

EXAMPLE   Eye glance duration and vehicle speed.

3.9
performance
result of skill application demonstrated by a participant in performing a driving task or Transport Information and Control System (TICS) related task

3.10
primary task
task to be calibrated

Note 1 to entry: The Lane Change Test in ISO 26022 is one example a task that simulates or approximates a primary driving task.

Note 2 to entry: This document considers the Detection Response Task (ISO 17488) is the primary task to be calibrated even if performed with or without driving.

3.11
primary driving task
activities that the driver must undertake while driving including navigating, path following, manoeuvring, avoiding obstacles, and controlling speed; and which a participant may perform through the duration of a test (simulated substitute for driving)

3.12
secondary task
non-driving related additional task

Note 1 to entry: A calibration task for the purpose of this document.

3.13
secondary task demand
sum of visual, auditory, cognitive, motor, and speech resource demands required by a non-driving related task

3.14
system paced secondary task
activity in which the change from the current to the next state in the interaction between user and system is initiated by the system

Note 1 to entry: The pace can be fixed or variable.

3.15
target bar
moving line on the critical tracking task display which indicates the task error

3.16
task
process of achieving a specific and measureable goal using a prescribed method

3.17
user paced secondary task
activity in which the change from the current to the next state in the interaction between user and system is initiated by the user

4   Abbreviated terms

CI   Confidence Interval

CTT   Critical Tracking Task

DRT   Detection Response Task

LCT   Lane Change Test

MDEV   Mean Deviation (According to ISO 26022)

SURT   Surrogate Reference Task

TICS   Transport Information and Control System[a]

[a] A list of TICS fundamental services has been defined by ISO/TC 204/WG 1.

5 Calibration tasks

5.1 Principle and overview

For calibration purposes, a standardized calibration task shall be used as a secondary task in a dual task setting in combination with a method to assess attentional demand due to the use of an in-vehicle system. The dual task setting shall include a primary task and the secondary calibration task.

Examples for driving-like dual task settings may include the operation of a TICS secondary task in the Lane Change Test (LCT), in a driving simulator environment, or the Detection Response Task (DRT).

The calibration shall be performed in a setting that is intended for the assessment of secondary tasks and follow the training and experimental procedures of that method for assessing attentional demand.

Development of the calibration tasks and associated procedures have used the ISO 26022 LCT, and ISO 17488 DRT to represent the primary task. While the calibration tasks described herein are intended to be applicable to other primary task implementations and dual task settings, care shall be taken to ensure that the conditions are sufficiently similar to those of ISO 26022 or ISO 17488 considering equipment and instructions to ensure a valid application of this document and its procedures.

5.2 Types of calibration tasks

There are various possibilities to realize a calibration task. In the following subclauses, three example alternatives are specified in detail. These alternatives include a system-paced secondary task (critical tracking task, CTT), a user-paced secondary task (surrogate reference task, SURT), and an auditory-vocal delayed digit recall task (n-back). Both the CTT and SURT alternatives represent visual-manual tasks that can be used in a dual task setting and are recommended as calibration tasks while using the LCT as the primary task. Whereas the n-back is a cognitive task that can be used in a dual task setting and is recommended as a calibration task while using the DRT as the primary task.

5.3 Critical Tracking Task (CTT)

5.3.1 Description

The CTT is a visual-manual task, which requires continuous control activity by the participant.

The participant controls the position of a vertically moving target bar with respect to a fiducial line (centreline) within a target area by manipulating up and down arrow keys. The arrow key control gives discrete commands to the target bar which moves it up and down. The up key moves the target bar up, and the down key moves it down.

The dynamics of the motion of the target bar are a first order instability. If the participant does nothing, the target bar moves (divergently) towards the edge of the display. The participant then has to make suitable corrective arrow key inputs to bring the target bar back towards the centre (the red dashed line in Figure 1).

A control system block diagram of the CTT is shown in Figure 2.
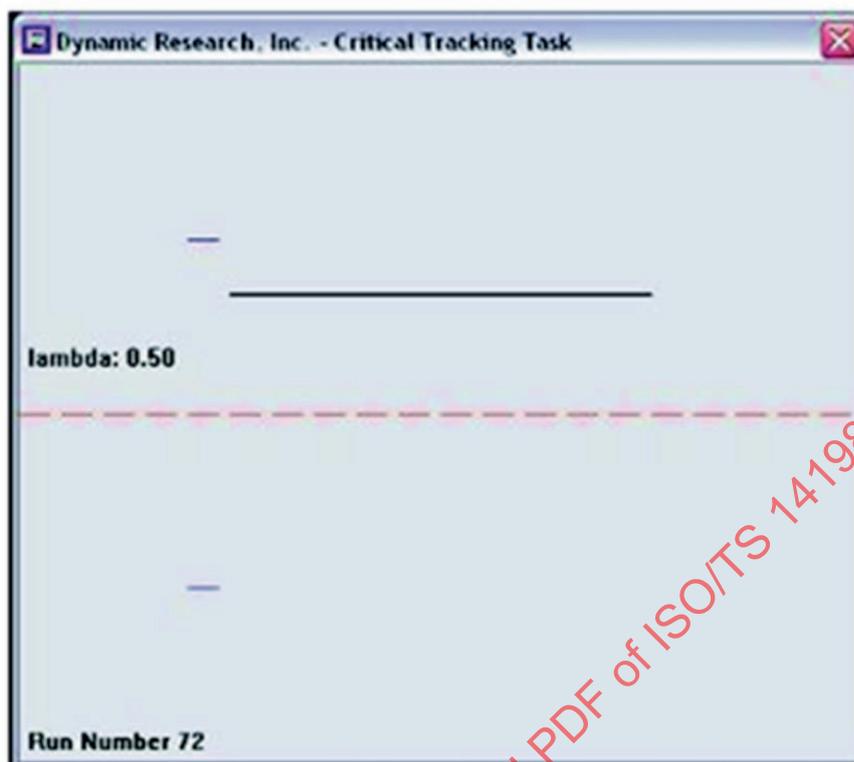
Key

1   target bar

2   centreline

Figure 1 — Typical screen of the CTT with target bar above the centreline

Key

λ Level of instability or rate of divergence (adjustable).

s Laplace transform variable.

Figure 2 — Control system block diagram of the CTT

5.3.2   Operation of the CTT

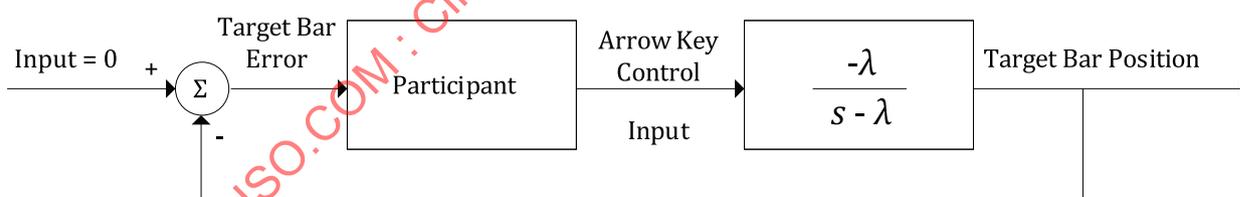When the program is first started, the screen will look like Figure 1. Nothing will happen initially. The centreline is displayed as a red dashed horizontal line in the centre of the display. The target bar is displayed as a black line. The target bar will start to move away from the centreline showing an increasing error. Two short blue reference lines are shown above and below the centre to subdivide the screen for better orientation. The instruction to the participant is to control the position of the target bar with the arrow keys (arrow up key and arrow down key) and keep it as close to the centreline as possible so as to minimize the errors.

5.3.3 Setup for CTT

The setup consists of a screen with SVGA resolution plus keyboard should be as follows:

— The subtended vision angle (width) of the display area relative to participant's eyes shall be 13° ± 1° horizontally.

— 19 inch (483 mm) screen for a width to height ratio of display area of 4:3. For a display with a width to height ratio of 16:9 (and equivalent to a 19 inch screen), one should use a setup of with a 23 inch (584 mm) screen.

— The centre of the secondary display shall be positioned 28° ± 2° horizontally (right or left depending on the intended display position in the vehicle, LHD or RHD) and 20° ± 2° vertically from participant's straight ahead line of sight. For further details see Annex A.

— To control the target bar movement, the arrow key of a standard PC keyboard (or an equivalent arrangement of keys, for example Figure 3) shall be used. Participants are allowed to place the keyboard in a comfortable position on the same side of the steering wheel as the CTT display on a table or console aside but not connected to their body.

Figure 3 — Example of key-pad to move the highlight and confirm target location result

5.3.4 Test conditions for CTT

The test in the dual task setting shall include one specified level of λ, 0,5, the CTT easy condition. The λ level shall be set prior to each run and kept constant during the run. In each run the primary task and the CTT task shall be performed for at least 2 min.

5.3.5 Participants for CTT

The CTT as a calibration task shall be performed in a dual task setting by at least 16 [n = 16] participants who are licensed drivers. Participants should be familiar with the primary task as well as operation of the CTT. The level of participant familiarity shall be documented in the protocol. In the case of calibration using an LCT setting, it is recommended to select the participants according to ISO 26022, and following the sample description regarding age, gender and familiarity with primary and secondary task in Bengler, K., Mattes, S., Hamm, O., Hensel, M., 2010[2]. Typically, this was an average age of 32 to 45, gender balanced, and instructed and practiced in both the primary and secondary tasks.

5.3.6 Participant instruction for CTT

The following verbal/written instructions for CTT shall be given to the participant:

"A horizontal black target bar is displayed on the secondary task screen. There is also a red dashed horizontal line in the centre of the screen. When the task is started, the black target bar will tend to

move away from the red centreline. The motion of the black target bar can be controlled by the arrow-up- and arrow-down keys in order to bring the target bar back to the centre. These keys give discrete commands to the target. (The down arrow key moves the target bar down; the up arrow key moves the target bar up.) The goal is to control the motion of the black target bar to keep it as close as possible to the red dashed centre line."

The λ level is set to a fixed value at the beginning of a run and defines the secondary task difficulty.

If the participant does nothing, the target bar moves towards the edge of the display. The further the black target bar is away from the centre line, the faster it will move, so a larger (longer duration) input is needed, to bring it back to the centreline.

The difficulty of controlling the black target bar can vary with the size of the error. The participant will need to work quickly and correctly in order to keep the target bar near the centreline.

5.3.7   Practice trials

Participants shall practice the CTT secondary task alone at λ levels of 0,5 and 1,0. Before proceeding to the dual task conditions, participants shall be able to perform the CTT alone at a level of $\lambda = 1,0$ for one minute successfully (i.e. without the target bar hitting the upper or lower limit of the display). Only the λ value of 0,5 is used in the dual task trials.

Primary task and dual task practice trials shall be conducted as prescribed in the corresponding standard or other procedures.

5.3.8   Test metrics

The primary task measures shall be those prescribed for the primary task. If the Lane Change Test is used, these are in ISO 26022.

Participants' performance in the CTT is measured using:

— CTT root mean square deviation of the target bar from the centreline in pixels.

— CTT percentage of time with target bar at the upper or lower limit.

## 5.4   Surrogate Reference Task (SURT)

5.4.1   Description

The Surrogate Reference Task (SURT) is both visually and manually demanding. The level of demand can be varied.

The SURT is a visual search paradigm. In a typical visual search paradigm, participants are asked to report whether or not a pre-specified target is embedded in a multi-item display, usually comprised of an array of alphanumeric symbols, forms, colours or words. The non-target items are called distractors. To vary the time the participant will need to identify the target and to answer (reaction time), the similarity between target and distractors can be manipulated. The more similar the distractors are to the target the more time is typically needed to identify the target. Unless these distractors have no unique feature distinguishing them from the target (feature with a visual "pop out" character), the number of distractors can also increase the time participants will need for search (Treisman and Gelade, 1980[36]). An example for the task is shown in Figure 4.

For the SURT, participants shall locate a pre-specified target circle which is either larger or smaller than the distractor circles (i.e., the visual component). Participants then shall manually select the region in which the target circle is located (i.e., the manual component). The distractor circles are of a uniform size compared to the target circle, which is distinguishable from the distractors based on its size. The diameter of the target remains consistent over the course of the trials. The thickness of the circle shall be equal for both the target and distractors.

For the participant to be able to indicate the position of the target, the display is divided into evenly distributed vertically arranged rectangular areas. During search none of these areas is marked and no highlight is visible. Once the participant visually locates the target, the participant should press the left or right key (indicated by graphics on the keys) depending on the position of the target on the display, and the highlight (as in Figure 4) will appear immediately to the left or right of the centreline of the display. Consecutive presses on the keys moves the highlight to the left or right. The participant uses the keys to move the highlight until it is positioned over the target. Finally, the participant shall confirm completion of the task by pressing a third key (confirmation key).

The response time is defined as the time from stimulus onset to the keypress of the confirmation key. The task is participant paced, i.e., after confirmation the next stimulus will appear without any time out. If the target is within the highlighted area when the confirmation key is pressed, the answer is recorded as correct, otherwise as incorrect. Response times are measured to estimate the participant's performance with different difficulty levels of the calibration task, and to confirm increasing calibration task difficulty.
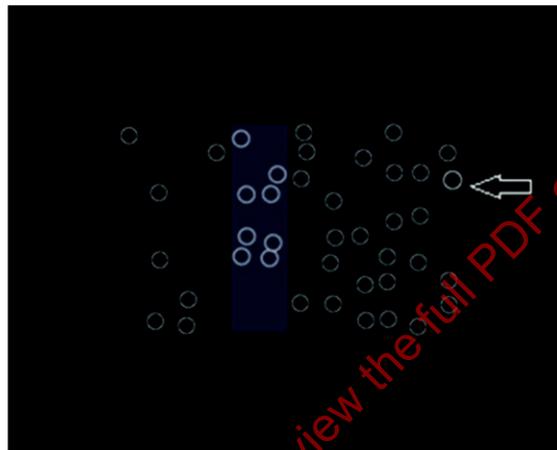
NOTE   Here the target is the right circle in the upper area marked by the arrow. The visual display of SURT in operation does not include the marking arrow

Figure 4 — Visual display of SURT distractors and target setup for SURT

5.4.2   Setup for SURT

Calibration task display setup should be as follows, in order to make all values consistent between tests or sites:

— Vision angle of the display area from participant's eyes shall be 13° ± 1° horizontally;

— 19 inch (483 mm) screen for a width to height ratio of display area of 4:3. Or 23 inch (584 mm) for a width to height ratio of display area of 16:9;

— Centre of display shall be positioned 28° ± 2° horizontally (right or left depending expected position of system to be tested) and 20° ± 2° vertically from participant's straight-ahead line of sight;

— Line thickness of target and distractors: 0,07° corresponding to approximately 1 mm when viewed at a distance of 800 mm;

— Target size: 0,7° corresponding to approximately 10 mm when viewed at a distance of 800 mm;

— Background colour: black (RGB-Values: 0, 0, 0);

— Target and distractor colours: light grey (RGB-Value: 192, 192, 192);

— Highlighting colour: medium grey (RGB-Value: 160, 160, 160);

— Minimum contrast (light/dark), target and distractors to background: 8:1 (ISO 15008).

An example setup is shown in Annex A.

The stimuli for targets and distractors are circles. Only the size of the circles distinguishes between target and distractors, while the size of the target remains constant (see Table 1). The three (3) difficulties of the SuRT include:

— Target is easy to find.

— Target is moderate to find.

— Target is hard to find.

Table 1 — Calibration task setup

Calibration task difficulty level	Number of distractors	Size of distractors	Number of rectangular areas
Easy	50	0,36°	2
Moderate	50	0,57°	6
Hard	50	0,64 °	8

Once the target circle is located, the manual answer is given using the arrow key of a standard PC keyboard or an equivalent keypad. The arrangement of the keys shown in Figure 3 may be used. The left/right keys move the shaded area over the target circle. Either up/down key can be used to enter the answer. Participants can place the keyboard at their own comfort on the same side of the steering wheel as the Calibration task display. Alternatively, the keyboard may be part of the Calibration task display unit (lap-top computer).

The background colour of the Calibration task display is black (RGB-Values: 0, 0, 0). Each display consists of one target and 50 distractors. Targets and distractors are randomly distributed over the whole display. The items should not intersect with each other.

To control the visual demand of the Calibration task it is possible to manipulate the size of the distractors (e.g. diameter). Distractor size closer to target size will result in higher visual demand. The more the distractor size differs from the target size the less visual demanding the Calibration task will be.

5.4.3 Test conditions for SURT

The test shall include one level of difficulty. The test duration should be approximately 2 min.

Figure 4 shows an example for the SURT hard condition to be used. Using a display distance of 78,5 cm, – the target size is 61 arcmin (14 mm); the distractor 53 arcmin (12 mm).

5.4.4 Participants for SURT

The SURT as a calibration task shall be performed in a dual task setting by at least 16 participants [n = 16]. Participants shall be familiar with the primary task as well as the SURT. The level of participants' familiarity with the primary task and the SURT shall be documented in the protocol. For example, by stating that the participant performed this secondary task previously in a dual task condition. In the case of calibration using the LCT procedure, it is recommended to select the participants following the protocols in ISO 26022 and the sample description in Bengler, K., Mattes, S., Hamm, O., Hensel, M., 2010[2]. For LCT being the primary task, the number of n = 16 participants is based on the experience of Bengler et al. 2010[2].

5.4.5 Participant instructions for SURT

All participants should be instructed the same way as follows:

An array of circles is displayed on the secondary task screen which is down and to the right of the driving simulation scene. One of these circles is bigger than the other ones. The position of this circle is randomly varying, but the positioning should not influence the difficulty to detect the difference. The participant task is to find the bigger circle and decide whether it is located in the left or in the right part of the screen. Accordingly, the participant has to press the left or right arrow key. Sometimes the decision is just left or right to highlight the bigger circle and sometimes the participant has to press the left or right key several times to highlight the bigger circle depending on the width of the highlighted area. To confirm the decision the participant has to press the [arrow-up/confirmation] key. The next trial starts automatically without any delay.

Instructions on dual task attention allocation should be those for the dual task procedure being used. In the case of the LCT procedure as one realization of a primary task, those instructions are given in ISO 26022:2010, 3.6 and Annex A.

5.4.6 Practice trials

Primary task and dual task practice trials shall be conducted as prescribed in the primary task settings and procedures.

Participants shall be able to perform the so called "hard" level of the SURT under a single task condition before going to the dual task condition.

5.4.7 Test metrics

The primary task performance measures shall be those prescribed for the primary task. If the Lane Change Test is used, these are in ISO 26022.

Participants' performance in the SURT is measured using:

— Percentage of correctly solved screens;

— Mean response time per SURT screen in seconds.

5.5 N-back

5.5.1 Description

The n-back is most frequently an auditory-vocal task where a participant listens to a continuous sequence of numbers and responds with the appropriate number back in the sequence. Additional versions of the n-back with a combination of visual, auditory, manual, and verbal have also been used.

An example of the n-back is, during a "1-back" if the numbers 5, 7, and 3 are presented through audio, a participant would say nothing when 5 is presented, 5 when 7 is presented, and 7 when 3 is presented. Table 2 shows examples of 0-back, 1-back, and 2-back using this example; these are the most common versions used.

Table 2 — Examples of the 0-back, 1-back, and 2-back tasks

0-back						1-back						2-back				
Presented	2	6	7	1		Presented	2	6	7	1		Presented	2	6	7	1
Response	2	6	7	1		Response	–	2	6	7		Response	–	–	2	6

5.5.2 Operation of the n-back

The n-back is a simple auditory-vocal task where an audio file is played with numbers being presented at a fixed interval. The participant shall respond during the intervals depending on which version of the n-back is being used. Participants are instructed to perform the task as accurately as possible

with the experimenter recording response accuracy. The task is started by the experimenter when the participant is prepared to perform the task.

The n-back task can be used in a dual task setting with the DRT as the primary task. Example setups for the DRT are shown in Annex C.

For the calibration task purpose, the DRT shall not be used in a non-driving set up. The DRT shall be used in combination with a driving task. The driving task can be either a surrogate-driving task, a simulator-driving task or an on-road driving task.

5.5.3 Setup for n-back

The n-back is presented through speakers or headphones. Care should be taken that the volume of the n-back task is sufficiently higher than other tasks being performed in order for the participant to hear it. The numbers are presented every 2,25 s from the beginning of the stimulus presentation (Mehler & Reimer, 2013[24]). The total amount of time that the n-back occurs can be dependent on the research question being answer. A typical time used for the n-back is 2 min ± 1 min.

5.5.4 Test conditions for n-back

The test in a dual task setting may include any of the versions of the n-back task. For comparison purposes, use of the 0-back and 1-back is encouraged. N-back versions should be counterbalanced in order to prevent learning effects when performing multiple tasks. There should be no additional audio cues between sequences within the n-back stream.

5.5.5 Participants for n-back

The n-back as a calibration task shall be performed in a dual task setting by at least 16 [n = 16] participants who are licensed drivers. There are no specific ages, gender, or task familiarity requirements but these variables should be recorded.

5.5.6 Participant instruction for n-back

The following is an example of verbal/written instructions for the n-back to be given to the participants (Mehler, Reimer, & Dusek, 2011[26]):

0-back (see Table 3): "The first version of the task is called the 0-back task, which simply means, that as I read each list of ten numbers, you are to repeat out loud the last number that you've heard. For example, if I were to say the number three, you would say three; then if I said two, you would say two; then if I said six, you would say six, and so on. Try to be as accurate as you can be."

Table 3 — 0-back

Presented:	3	2	6	7	1
Response:	3	2	6	7	1

1-back (see Table 4): "The second version of the task is called the 1-back task, which simply means that as I read each list of ten numbers, you are to repeat out loud the number before the last number that you heard. For example, if I said 3, you would say nothing, then if I said 2, you would say 3, then if I said 6, you would say 2, and so on. Try to be as accurate as you can be."

Table 4 — 1-back

Presented:	3	2	6	7	1
Response:		3	2	6	7

2-back (see Table 5): "The final version of the task is called the 2-back task, which simply means that as I read each list of ten numbers, you are to repeat out loud the number that was read two numbers ago.

For example, if I were to say the number 3, you would say nothing, then if I said the number 2, you would say nothing, then if I said 6, you would say 3, if I say 7, you would say 2, and so on. Try to be as accurate as you can be."

Table 5 — 2-back

Presented:	3	2	6	7	1
Response:			3	2	6

5.5.7 Practice trials

Participants shall practice the n-back task alone prior to performing it in dual task conditions. Performance criteria should adhere to ISO 17488 to determine if participants are comfortable with the task and capable of completing it in dual task conditions.

Primary task and dual task practice trials shall be conducted as prescribed in the corresponding standard or other procedures.

5.5.8 Test metrics

The primary task measures shall be those prescribed for the primary task.

Participants' performance in the n-back is measured by calculating response accuracy (e.g., number of accurate responses divided by total responses).

6 Calibration criterion

6.1 Calibration criterion procedure

The calibration assessment shall be performed using a primary task performance measure, and confidence intervals for the respective metric. The performance values in the primary task and the secondary tasks of the laboratory under calibration shall be compared to a confidence interval based on the data provided in Annexes B and D as a reference zone of indifference.
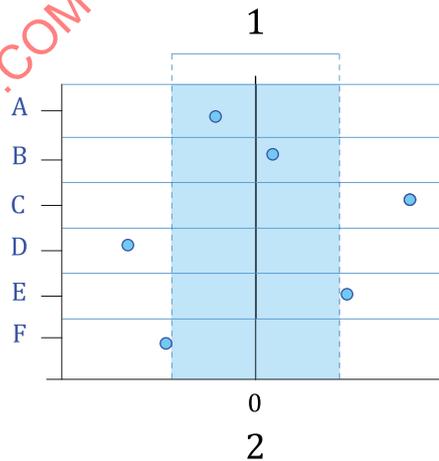
Key

1 99 % CI
2 treatment effect

Figure 5 — Possible outcomes for six different laboratory values (A to F) and zone of indifference calibration data

Figure 5 illustrates how the performance values of a laboratory under calibration might correspond to a calibration data set.

A laboratory setting shall be called calibrated if the values of the test sample are within a one-sided 99 % CI (or two-sided 98 % CI) of the calibration data set for their corresponding values.

For the purpose of calibration, case A or B is desired, whereas cases C, D are clear indications of an absence of equivalence. Cases E and F are acceptable for calibration, but it is strongly recommended to check the data set and laboratory situation.

If the E or F case occurs:

— In the case of LCT, it shall be checked using the primary task performance measure in the baseline single task condition, and the SURT hard and CTT (λ level = 0,5) dual task conditions. At the same time, the secondary task performance measure should be at least at a minimum level for the CTT and for the SURT (see Annex B values from LCT multi-lab data set).

— In the case of DRT, it shall be checked using the DRT performance measure in the baseline task condition, and the n-back (0 and 1) dual task conditions. At the same time, the secondary task performance measure at the n-back task should be checked and participants who do not succeed in doing it should be excluded.

The level of calibration shall be checked in three conditions following a staged approach. It is recommended to check the value of the baseline performance of each subject in the primary task in advance of the further calibration analysis. The following conditions are:

Conditions 1, 2 and 3 in the case of LCT:

— Compare the primary driving task baseline performance mean value of the sample to the 99 % CI of the corresponding reference data set (given for the LCT in Table B.1);

— Compare the primary driving task performance mean value in the CTT dual task condition to the 99 % CI of the multi-lab data set, and compare the CTT performances (given for the LCT in Tables B.1 and B.3);

— Compare the primary driving task performance mean value in the SURT dual task condition to the 99 % CI of the multi-lab data set, and compare the SURT performances (given for the LCT in Tables B.1 and B.2).

Conditions 1, 2 and 3 in the case of DRT:

— Compare the RT values for DRT in the baseline condition of the sample to the 99 % CI of the corresponding reference data set (given in Table D.1);

— Compare the RT values for DRT in the 0-back task condition to the 99 % CI of the multi-lab data set (given in Table D.1);

— Compare the RT values for DRT in the 1-back task condition to the 99 % CI of the multi-lab data set (given in Table D.1).

A laboratory shall be called:

— (I) fully calibrated, if in conditions 1, 2 and 3 are satisfied;

— (II) partially calibrated, if only conditions 1 and 2 or conditions 1 and 3 are satisfied;

— (III) not calibrated, if only condition 1 is satisfied.

For a more sophisticated description of methods for equivalence testing refer to Wellek, S. (2003) [37].

6.2   General calibration considerations

In general, the following conditions shall be satisfied as part of the calibration procedure:

In the case of LCT:

— Performance in baseline condition is better than performance in the dual task conditions (SURT, CTT, or n-back conditions);

— In the case of LCT: Primary task performance in a dual task, for example, CTT_easy condition is better than primary task performance in a dual task SURT_hard condition;

— In the case of DRT: the RT values for DRT in 0-back condition are lower than the RT values for DRT in 1-back condition.

In the event that the calibration criteria and considerations are not satisfied, the experimenter should check that data analysis followed the prescribed procedures in this document and the prescription of the primary task. Further, all equipment and setup should be checked for consistency with the prescriptions in this document. For example, different results can be obtained if an experiment is conducted on a desktop computer or using a more extensive driving simulator (Bruyas et al. 2008[5]). This may be due to differences in the simulated vehicle dynamics, the steering feel characteristics, the size and quality of the primary task roadway image, etc. In the case of visual-manual tasks, the screen position may also influence the results as longer RT are generally registered when the display installation is of lower vertical eccentricity (Koustanai et al. 2018[23]).

A general linear shift of mean values above or below the levels shown in the Annex B and D reference tables for the primary task performance can be due to a selection of participants as weak or strong in task performance. There is evidence in the data of Bengler et al. 2010[2] that age and experience with the primary task and/or with the secondary task can be moderating factors.

Especially with regard to II and III, it is recommended to check the secondary task equipment, and instructions on dual task conditions.

The calibration procedure may be repeated after a suspected source for the lack of calibration has been removed.

Despite using the specified procedures and due scientific care, it may be the case that a laboratory is still unable to achieve "calibrated" status in accordance with this document. This may not be the result of a scientific or technical deficiency on the part of the laboratory. Instead, it may reflect differences in the participants' understanding and interpretation of instructions and procedures on the part of the participants, or the effect of other local customs regarding driver behaviour or attentional factors in general.

Annex A
(informative)

Calibration task setup details using the LCT

A typical experimental setup involves a monitor for the primary task, and a 19 inch (483 mm) diagonal screen monitor (width to height ratio of 4:3) for the Calibration task display. The participant should sit directly in front of the primary task monitor. In the example of Figure A.1 an LCT primary task is presented on a monitor at a distance of 900 mm (viewing angles of 0° horizontally and 0° vertically to centre of monitor). This results in an angle of vision of approximately 20° horizontally for the primary task (LCT) scene on a 16 inch (406 mm) diagonal monitor.

The following values for the Calibration task set up will then apply:

— Calibration task display position relative to LCT monitor: 240 mm downward, 350 mm rightward, 240 mm forward (for a left-hand drive country).

— Resulting distance to Calibration task display: 785 mm.

— Resulting viewing angle to Calibration task display: 28° horizontally (rightwards) and 20° vertically (downwards).

— Resulting angle of vision: 13° ± 1° horizontally.

In countries where a right-hand drive vehicle is used, the Calibration task screen should be placed left-downward to the visual scene of an LCT-Simulation.

The experimental setup is shown in Figure A.1 for a left-hand drive country. A 16 inch monitor can be used for the visual scene of an LCT-simulation primary task. A steering wheel is used for the LCT-simulation. The display for the Calibration task is on the lower right in the front view. A photograph is shown in Figure A.2.
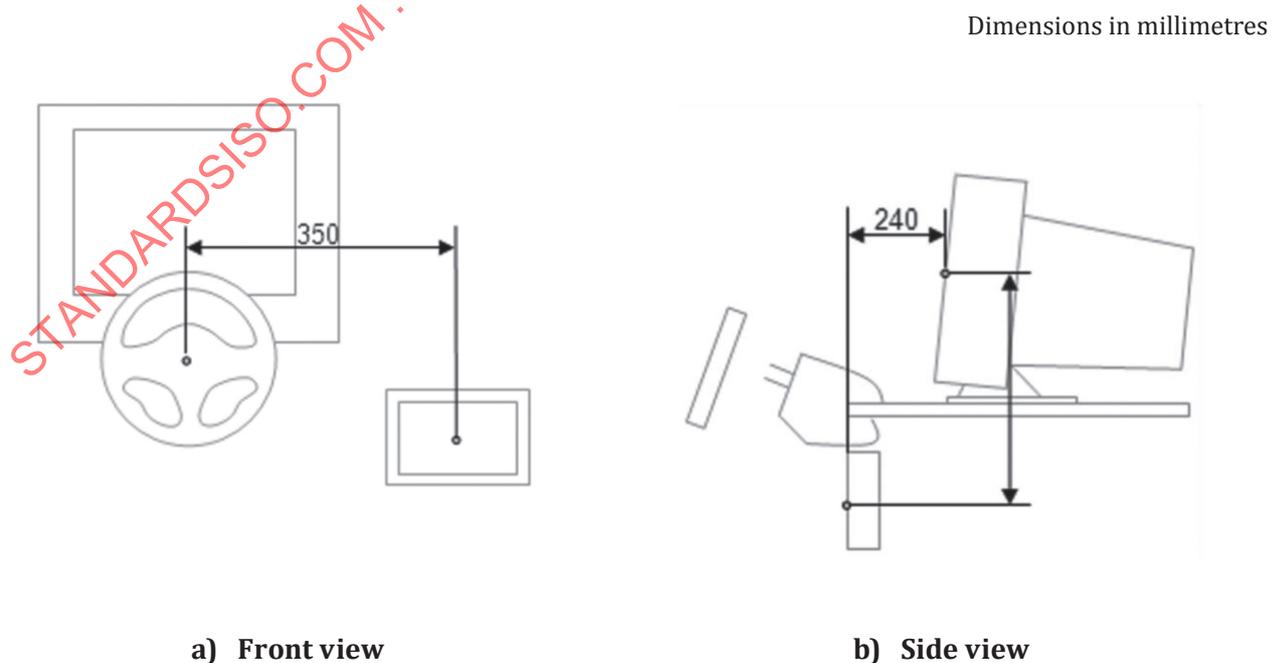
Dimensions in millimetres

a) **Front view**   b) **Side view**

Figure A.1 — Example setup for primary task (LCT example) and secondary task display
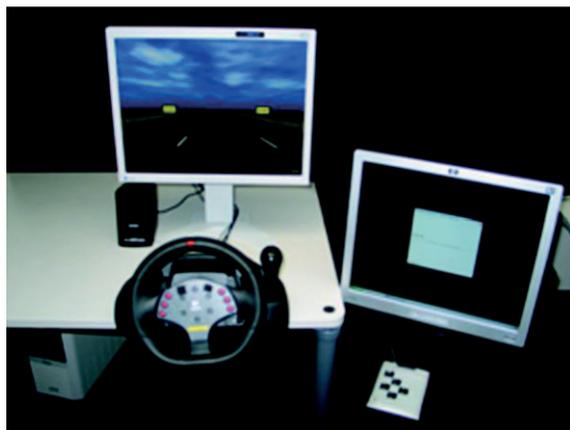
Figure A.2 — Setup for primary task (LCT example) and secondary task (CTT example)

Annex B
(informative)

Multi-lab reference data for LCT

The following tables give an overview of comparison values for the assessment of calibration criterion using the LCT as the primary task (see Clause 6). The values are calculated on the basis of the Multi-Laboratory Calibration Study (Bengler, K., Mattes, S., Hamm, O., Hensel, M., 2010[2]). This study used the Lane Change Test as the primary task.

If the LCT is being used as the primary driving task the level of calibration should be checked in three conditions following a staged approach. If the LCT is used as the primary task it is recommended to check the basic $MDEV_{baseline}$ value of each subject in advance of the further calibration analysis. The basic $MDEV_{baseline}$ value should not exceed a level of 1,2 in accordance with the ISO 26022 training procedure. The conditions are:

— Compare LCT adaptive $MDEV_{baseline}$ mean value to the 99 % CI of the multi-lab data-set (given by Table B.1);

— Compare LCT adaptive $MDEV_{CTT}$ to the 99 % CI of the multi-lab data-set, and compare CTT performance (given by Tables B.1 and B.3);

— Compare LCT adaptive $MDEV_{SURT}$ to the 99 % CI of the multi-lab data-set, and compare SURT performance (given by Tables B.1 and B.2).

With LCT as the primary task, a laboratory should be called:

— (I) fully calibrated, if in conditions 1, 2 and 3 the laboratory adaptive $MDEV_{baseline}$, $MDEV_{SURT}$ and $MDEV_{CTT}$ values lie within the corresponding confidence intervals;

— (II) partially calibrated, if in conditions 1 and 2 or 1 and 3 the laboratory adaptive $MDEV_{baseline}$ and $MDEV_{SURT}$ and/or $MDEV_{CTT}$ lie within the corresponding confidence intervals;

— (III) not calibrated, if in condition 1 only the laboratory primary task adaptive $MDEV_{baseline}$ lies within the corresponding confidence intervals.

Table B.1 — Adaptive MDEV values for LCT: Baseline and with CTT and SURT secondary tasks

Tasks	Number of labs	Mean	StdDev	Stder	99 % -CI Lower	99 % -CI Upper
Baseline	7	0,37	0,03	0,01	0,32	0,42
CTT easy	6	0,60	0,04	0,01	0,54	0,66
SURT hard	6	0,73	0,05	0,02	0,64	0,82

Table B.2 — Secondary task performance SuRT response time in the hard condition

	Number of labs	Mean [sec]	StdDev	Stder	99 % -CI Lower	99 % -CI Upper
Response Time	6	7,29	1,22	0,50	5,28	9,30

Table B.3 — Secondary task performance CTT root mean square deviation in the $\lambda = 0{,}5$ condition

	Number of labs	Mean	StdDev	Stder	99 % -CI Lower	99 % -CI Upper
Deviation	6	7,56	2,19	0,89	3,96	11,16

Some feedback from the Interlab experiment could be helpful in the event that a laboratory appears not calibrated. It has been shown that:

— if LCT is used as the primary task, the experimenter has to ensure that basic mdev values follow the requirement not to exceed the criterion value of 1,2 as formulated in ISO 26022;

— drivers with too much experience in either the primary task or a secondary task may cause a sample to be outside the 99 % CI (too low values);

— a more complex (incorrect) version of a calibration task could lead to poorer driver performance (too high values), and

— also, sometimes wrong calculations can give wrong results.

The mean age of the sample participants could also influence the data; older participants may obtain poorer results (higher values).

Annex C
(informative)

Calibration task setup details using the DRT

The DRT method is based on a simple detection-response task where participants respond to relatively frequent artificial stimuli presented with a specified degree of temporal uncertainty. Detection performance, measured in terms of response time and hit rate, is assumed to represent the degree to which attention is affected by the demand and, in particular, the cognitive load component imposed by the secondary task under evaluation. Longer reaction times and reduced hit rate are indicative of higher cognitive load (ISO 17488).

The DRT can be used in different experimental setups:

— In a **non-driving experimental setup**, the DRT is performed concurrently with the secondary task under evaluation in a non-driving condition.

— In a **surrogate driving experimental setup**, the DRT is performed concurrently with the secondary task under evaluation <u>and</u> some primary task elements that function as a surrogate for driving. DRT performance with a secondary task under evaluation is compared to a baseline condition where the DRT is performed concurrently with the surrogate primary task.

— In a **driving simulator experimental setup**, the DRT, secondary task, and driving task are executed concurrently. In this experimental setup, DRT performance with a secondary task under evaluation is assessed relative to a baseline condition where the DRT is performed while driving without the secondary task.

— In an **on-road experimental setup**, the DRT, secondary task, and driving task are executed concurrently on either a closed track or an open road with real traffic. In this experimental setup, DRT performance with a secondary task under evaluation is assessed relative to a baseline condition where the DRT is performed while driving without the secondary task.