CEN WORKSHOP AGREEMENT

CWA 17663

February 2021

ICS 13.180

English version

Measurement of Worker Satisfaction in Automated Systems - Methodology CEN Workshop Agreement

This CEN Workshop Agreement has been drafted and approved by a Workshop of representatives of interested parties, the constitution of which is indicated in the foreword of this Workshop Agreement.

The formal process followed by the Workshop in the development of this Workshop Agreement has been endorsed by the National Members of CEN but neither the National Members of CEN nor the CEN-CENELEC Management Centre can be held accountable for the technical content of this CEN Workshop Agreement or possible conflicts with standards or legislation.

This CEN Workshop Agreement can in no way be held as being an official standard developed by CEN and its Members.

This CEN Workshop Agreement is publicly available as a reference document from the CEN Members National Standard Bodies.

CEN and CENELEC members are the national standards bodies and national electrotechnical committees of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Republic of North Macedonia, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and United Kingdom.

EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

CEN-CENELEC Management Centre: Rue de la Science 23, B-1040 Brussels

Ref. No.:CWA 17663:2021 E

Contents

European Foreword

CWA 17663:2021 has been developed in accordance with the CEN-CENELEC Guide 29 "CEN/CENELEC Workshop Agreements – A rapid prototyping to standardization" and with the relevant provisions of CEN/CENELEC Internal Regulations – Part 2. It was agreed on 2021-01-18 by a Workshop of representatives of interested parties, the constitution of which was supported by CEN following the public call for participation made on 2020-04-30. However, this CEN Workshop Agreement does not necessarily reflect the views of all stakeholders that might have an interest in its subject matter.

The final text of CWA 17663:2021 was provided to CEN for publication on 2021-02-12.

The following organizations and individuals developed and approved this CEN Workshop Agreement:

- CRANFIELD UNIVERSITY, UK, Dr. Sarah Fletcher – Chairperson
- UNE, Spain, Ms. Marta Fernández – Secretary
- CEA LIST, France, Ms. Margarita Anastassova
- CIAOTECH S.R.L. A SOCIO UNICO PNO INNOVATION B.V., Italy, Ms. Chiara Zocchi
- CRANFIELD UNIVERSITY, UK, Ms. Teegan Johnson
- FUNDACIÓN TEKNIKER, Spain, Mr. Jon Larreina
- HÉROUX-DEVTEK, Spain, Ms. María del Mar Otero
- I.S.A.R. KOMAT (INGENIERÍA Y SERVICIOS DE AUTOMATIZACIÓN Y ROBÓTICA), Spain, Mr. Iban Azurmendi
- LANTEGUI BATUAK, Spain, Mr. Miguel Martín
- PREVENCONTROL, Spain, Mr. Gustavo Rosal
- RWTH AACHEN UNIVERSITY (RWTH) LABORATORY FOR MACHINE TOOLS AND PRODUCTION ENGINEERING (WZL), Germany, Mr. Florian Becker

Attention is drawn to the possibility that some elements of CWA 17663:2021 may be subject to patent rights. The CEN-CENELEC policy on patent rights is described in CEN-CENELEC Guide 8 "Guidelines for Implementation of the Common IPR Policy on Patents". CEN shall not be held responsible for identifying any or all such patent rights.

The Workshop participants have made every effort to ensure the reliability and accuracy of the technical and non-technical content of CWA 17663:2021, but this does not guarantee, either explicitly or implicitly, its correctness. Users of CWA 17663:2021 should be aware that neither the Workshop participants, nor CEN can be held liable for damages or losses of any kind whatsoever which may arise from its application. Users of CWA 17663:2021 do so on their own responsibility and at their own risk. The CEN Workshop Agreement should not be construed as legal advice authoritatively endorsed by CEN/CENELEC

This CWA has been proposed by the A4BLUE European Project (funding from the European Commission's Horizon 2020 – The Framework Programme for Research and Innovation (2014 - 2020) under Grant Agreement GA 723828), whose main objective is the development and evaluation of work systems that are adaptive to deal with evolving requirements of manufacturing processes and human operators. It is proposed in order to fill a gap in existing standards for ergonomics and human-system interaction which deal with usability and its outcomes but do not specifically address the need for accurate and reliable measurement of worker satisfaction in automated work systems.

Introduction

The rapid development of automated work systems (3.1) that have increasingly responsive and self-adapting functions is transforming manual work. As people work ever more closely and collaboratively with automation, these functions will not only be designed to meet performance output requirements but also to satisfy the different preferences and capabilities of individuals. As the success of any system relies on user acceptance and engagement, worker responses should be a critical design consideration. Thus, system designers need to understand which and how system characteristics should be adjusted to optimise the outcomes of human-system interactions.

A key outcome of any system, product or service is the user's resultant state of 'satisfaction', defined as the *"extent to which the user's physical, cognitive and emotional responses that result from the use of a system, product or service meet the user's needs and expectations"* (ISO 9241-11:2018). Satisfaction is a latent psychological state/response that cannot be directly observed, similar to other internal cognitive/affective based outcomes like mental workload, stress, etc. User satisfaction is important to the efficacy of human-system interactions at work because numerous studies have demonstrated relationships between workforce satisfaction and productivity/performance as well as on longer term health and well-being outcomes, and have provided notable examples where new work processes have failed due to limited workforce approval and adoption rather than technical or functional issues. Indeed, user satisfaction is identified as one of the three principal outcome components of usability along with 'efficiency' and 'effectiveness' (ISO 9241-11:2018; ISO 6385:2016), as shown in Figure 1.

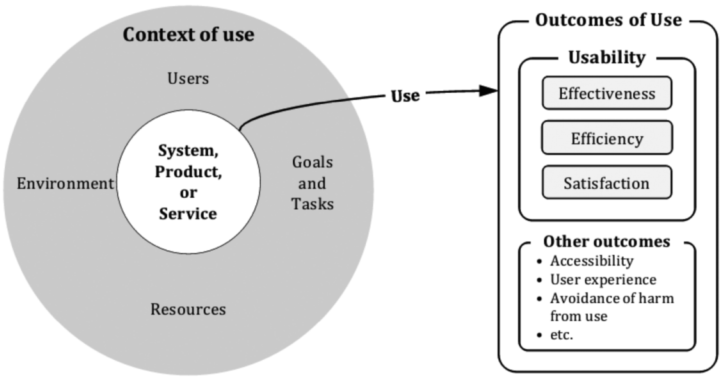

Figure 1 — Satisfaction as one of three 'outcomes of use' usability components (ISO 9241-11:2018)

Current standards advocate that user satisfaction is an important outcome that should be measured as part of ergonomic work system design, similarly to efficiency and effectiveness (ISO 6385:2016, Clause 4.5) but do not provide practical guidance on how to do so. Prescription of specific methods for measuring usability outcomes has been avoided because of the difficulty in addressing variations across contexts: *"[T]here is no single intrinsic measure of the usability of a system, product or service because effectiveness, efficiency and satisfaction depend on the users, goals and other components of the context of use* (3.2) *for which usability is being considered"* (ISO 9241-11:2018, 5.1). As a psychological state/response is an outcome produced by a number of different context-specific characteristics it will comprise a range of different 'dimensions'. For example, the innate characteristic of *extroversion* is just one dimension of personality, and itself comprises a number of sub -dimensions, e.g. assertiveness, sociability, etc. Similarly, a more temporal state or response like comfort is an outcome produced by the characteristics of a context being experienced at the time, so its dimensions will reflect responses to each of those. Thus, to get an overall measurement of a psychological state or responses it is important to measure all of the key dimensions that contribute to it in relation to the particular context requirements: *"components of satisfaction that are important will depend on the reasons for considering usability"* (ISO 9241-11:2018, 6.4.1). As a single tool would not fulfil the requirements of different systems, tailored/tailorable tools are needed.

'Psychometrics' (psychological measurement) is an approach that offers a robust methodology for developing tailored measures of psychological state or response dimensions, either for general applications or for application across specific contexts. Psychometric tests "attempt to analyse a person in terms of fundamental psychological characteristics" (Smith and Smith, 2005) by applying a "systematic procedure for observing behaviour and describing it with the aid of numerical scales or fixed categories" (Cronbach and Furby, 1970). This systematic procedure of scales or fixed categories will typically consist of multiple items (statements/questions) that have been carefully designed and iteratively tested with statistical procedures to ensure validity and reliability of the results. Multiple items are needed in order to interrogate and measure each of the different dimensions of the particular psychological state or response being investigated.

A number of psychometric tests already exist for measuring satisfaction in different contexts but there are currently no available/published tools for applications in the new context of adaptive automated work systems. Existing tests to measure 'job satisfaction' measure intrinsic factors (task-related factors, e.g. autonomy, creativity, etc) and extrinsic factors (factors external to the task, e.g. pay, conditions, etc). One set of job satisfaction measurement scales is specifically designed for industrial 'blue collar' workers' (Warr, Cook and Wall, 1979). However, no existing tools measure satisfaction in relation to aspects of new adaptive automation work systems. Thus, as user satisfaction may be influenced by a wide range of factors, and those that increase satisfaction are not necessarily the same as those that decrease dissatisfaction, system designers need a tool that will inform them of the particular characteristics of the systems they are designing that will increase or decrease satisfaction.

One particular methodology has been successfully applied to develop a valid and reliable psychometric (3.5) measure in the context of human-system interaction within automated work systems. In previous research conducted by Cranfield University a new measure of trust in industrial robots was created and tested for reliability and validity (Charalambous et al., 2016). As part of the A4BLUE project (Adaptive Automation in Assembly for BLUE collar workers' satisfaction in Evolvable contexts), this particular methodology was utilised towards development of a worker satisfaction measurement tool for the specific context of adaptive automated work systems. It offers a systematic framework for developing psychometric measurement tools to enable bespoke valid and reliable human outcome analysis across systems/contexts. For this reason the current document proposes that it is disseminated as part of a CWA for the measurement of worker satisfaction in automated work systems, which are becoming increasingly adaptive to human interaction.

Although efficiency and effectiveness are cited as key objectives for ergonomic work system design current standards – satisfaction is not (ISO 6385:2016, Clause 3.1). It is also not addressed by any designated standards in the same way as other user outcomes, e.g. mental workload (ISO 10075-1/-2/-3). This reflects that user satisfaction has not before been considered a critical outcome of system design. However, the closer and more collaborative interactions that will be necessary between workers and adaptive automation in future work systems will bring new impacts on safety and performance that are likely to involve satisfaction more critically. The purpose of this CWA is to provide guidance on how the psychometric development methodology applied by Cranfield University and the A4BLUE project can be used to develop bespoke, context-specific, valid and reliable psychometric tools for the measurement of user satisfaction for automated system design.

1 Scope

This CWA sets out guidance for the application of a systematic and reliable methodology which may be used to develop bespoke worker satisfaction (3.7) measurement tools for automated work systems (3.1) design. In doing so, it aims to promote the availability and consistency of robust psychometric (3.5) measurement tools for the design of future manufacturing systems in order to enhance worker satisfaction and, in turn, wider workforce wellbeing and performance outcomes. It does not advocate a single satisfaction measurement tool, because no single measure is universally applicable across different contexts.

The methodology described in this CWA document focuses on worker satisfaction measurement but, as it is based on social science principles for psychometric tool development, is transferable to the development of psychometric measures for measurement of other latent psychological variables (3.4) and other contexts.

The document offers a methodology for assessing psychosocial impacts of automation/human-robot cell design which is independent from risk assessment but could be used to support it. The methodology is not mandatory for a PSR-related workplace design or companies OSH-prevention policies.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

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

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

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

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

3.1
automated work system
a system comprising one or more workers and equipment that includes automated components that act together to perform a work task or goal

3.2
context of use
combination of users, goals and tasks, resources, and environment

3.3
item
an item is a statement or question that is posed to a test participant as part of a survey or questionnaire so that their response indicates their current subjective viewpoint, typically along a numerical scale

3.4
latent variable
unobservable or hidden variables, such as those manifest in human emotion and cognition

3.5
psychometric
quantitative measurement of psychological states, attributes, behaviour and performance via valid and reliable tests, questionnaires, and other measurement instruments

3.6
psychometric tests
valid and reliable survey-based assessment tools designed for measurement and analysis of psychological states, attributes, behaviour and performance

3.7
satisfaction
the extent to which the user's physical, cognitive and emotional responses that result from the use of a system, product or service meet the user’s needs and expectations

[SOURCE: ISO 9241-11:2018]

4 Methodology for the measurement of worker satisfaction in automated systems

4.1 General

The objective of providing a standard methodology for the measurement of worker satisfaction in automated work systems is to provide a means of reliable and harmonised assessment which will enable optimisation of satisfaction in the system design. There are variations in methods that can be used to develop psychometric tools and to establish evidence of psychometric properties of a tool, but the methodology provided in this CWA offers a systematic protocol for building valid and reliable context-specific measurement tools appropriate for automated work system design. There are three principal stages of the methodology: exploratory study, questionnaire design, and experimental study, each comprising a number of constituent steps, as shown in Figure 2, and as described in the following sections.

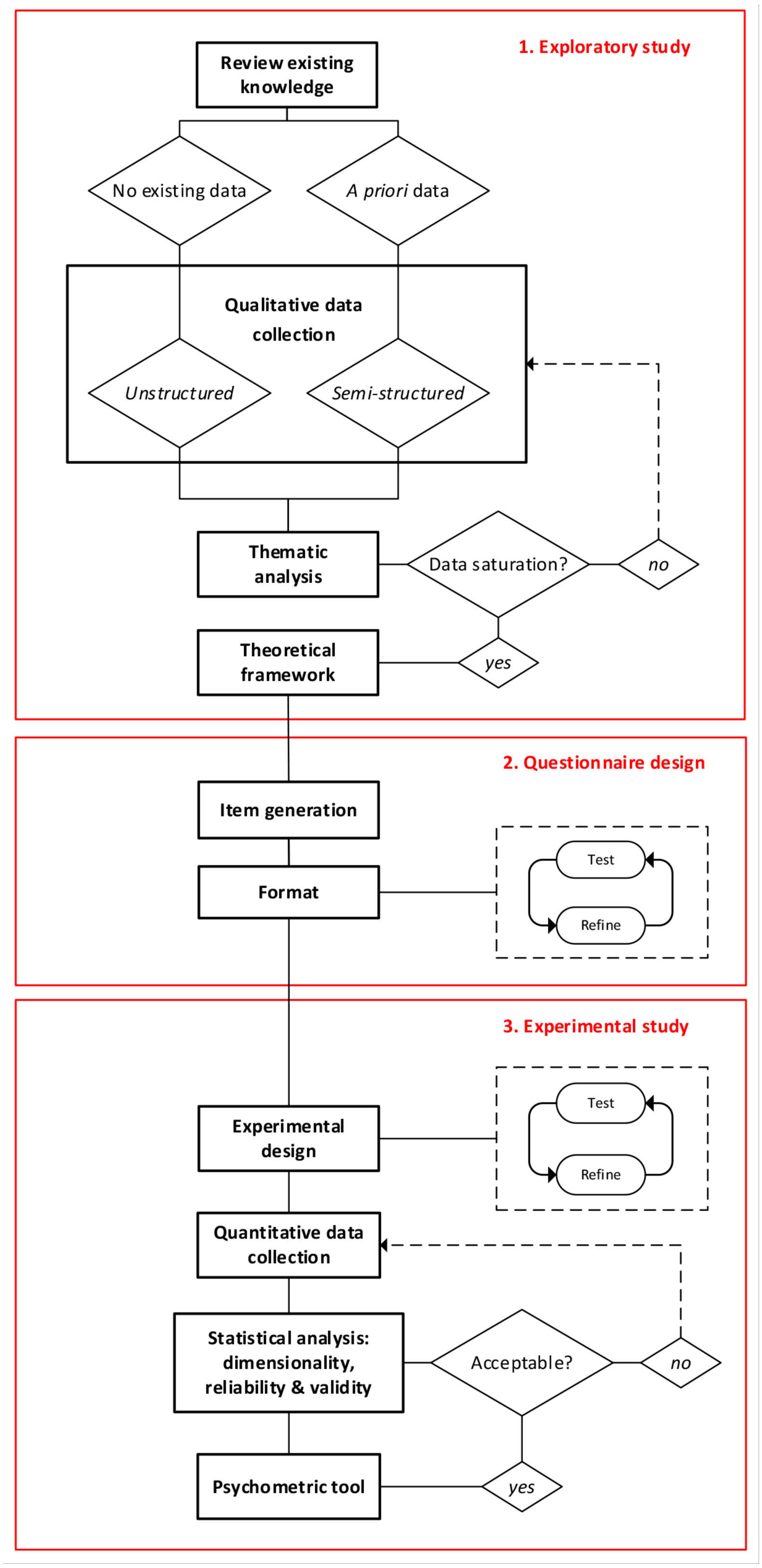

Figure 2 — Key steps of methodology for developing a psychometric measure

4.2 Exploratory study

4.2.1 General

Psychological states/responses consist of dimensions that are each induced and shaped by various situational and environmental characteristics. This means that, for the purposes of design, measurement of psychological states/responses should capture these individual dimensions in order to identify the characteristics that need to be optimised. In the development of a new tool for psychometric measurement for a specific context, it is necessary to generate new data that captures these key dimensions and characteristics. Thus, the first step of developing a tool to measure satisfaction in a new/undiscovered context, such as for the design of automated work systems, will involve exploratory study involving qualitative data collection. This will involve gathering data from participants to build up a picture of the specific psychological dimensions and context characteristics that need to be examined.

Research ethics protocols should be followed at all times in accordance with appropriate codes of practice for human participants, as set out by leading psychological associations, such as the British Psychological Society, American Psychological Association and European Federation of Psychologists Association.

Review existing knowledge

An exploratory study aims to explore and generate new knowledge about a particular context or phenomenon. Thus, an important first step is to conduct a broad review of existing relevant literature and practices to establish whether any information may already exist that could provide clues to help shape the investigation.

There are two principal outcomes that need to be established: a) *no existing data* can be found to indicate the dimensions and characteristics that are relevant to the context in question, or b) evidence is found of *a priori data*, i.e. evidence of potentially relevant factors (psychological dimensions and context characteristics).

4.2.2 Qualitative data collection

If the review of existing knowledge found *no existing data* to indicate dimensions and characteristics of relevance to the context, it is important to proceed without any prior assumptions. Thus, without existing reliable data an unstructured qualitative data collection method (interviews, focus groups, etc.) should be used, without a guiding framework of predetermined concepts. This will allow the data to build an entirely new framework without leading the participants, allowing them complete freedom to develop ideas and thoughts about the topic.

If *a priori data* is found in the review of existing knowledge, i.e. evidence of likely or potentially relevant factors (psychological dimensions and context characteristics) that need to be measured by the psychometric tool, these need to be explored to verify whether they are indeed relevant to the new context. This means the selected qualitative data collection method should involve a semi-structured framework whereby structure is formed by a planned line of inquiry to explore these topics, but it also enables an unstructured aspect so participants are still given freedom to also consider other emergent concepts and generate new data.

In both unstructured and semi-structured approaches, participants' responses should be recorded and subsequently evaluated to identify topics and themes it contains using the thematic analysis method described in section 4.2.3. Data collection should then continue until a point of 'saturation', i.e. when no new information is emerging from new data. At this point it can be assumed that the key factors in this context have been captured and, therefore, sampling (participant recruitment) can end.

Data collection should ideally involve a participant sample that reflects the most relevant wider population. Thus, the ideal sample for the context of work with automated work systems would be representatives of the workforce. However, to cover potential future changes in the given population it may be beneficial to employ a wider sample for inclusivity. For example, although a workforce may currently consist of a skewed bias towards gender/age/ability etc., it may be useful to consider whether this will change in the future and, if so, whether the characteristics of the exploratory study participants should be more varied. Participants should ideally have/be given experience of the context so that their subjective reports and discussions are based on direct understanding and familiarity, as opposed to imagined and indirect knowledge. Depending on the context under investigation, responses may be captured during participants' experience of it (continuously or at key points) or retrospectively, after the participants' experience has been completed.

4.2.3 Thematic analysis

As described, data collection and analysis of themes should continue until saturation and sampling targets have been reached. Thematic analysis, a form of content analysis, involves transcription of participants' comments, in full, followed by systematic scrutiny of the data to identify and categorise common topics. This process will codify the key characteristics for the given situation or context (dimensions). For this, Template Analysis is a recommended technique (King, 1998) as this develops a coding template in which key themes are organised hierarchically. Principal themes found in the text are first identified to form the top-level coding template and then any sub-themes are coded beneath. These themes and sub-themes essentially reflect the dimensions and sub-dimensions of the context that participants have reported. If a set of *a priori* questions have been used to ensure anticipated potential topics were covered in the data collection, the topics may be used as an initial template of themes to begin the analysis. However, the analyst must remain flexible as it may be appropriate to adapt the template in response to new topics that emerge; this is a 'hybrid approach' as use of *a priori* topics is deductive analysis but it could lead to coding of empirical data which would constitute an inductive approach.

The commonality of a topic or characteristic in participants' responses reflects that it is relevant to the specific situation or context being examined and, therefore, should be a candidate for the psychometric tool. Thus, if a topic or characteristic is mentioned by participants in relation to their satisfaction with a particular automated work system situation or context this indicates that it should be measured by the psychometric tool.

When all of the common topics in the data have been identified and coded into themes (and sub-themes) in the template until the point of saturation can be assumed, inter-rater reliability testing should then be conducted. To do this, other independent raters should apply the coding template to a selected portion of the data and their analysis should be compared by calculating the degree of concordance between their coding and the original coding. This should be verified using a suitable method, such as a Kappa or correlation coefficient, which shows level of rating concordance corrected for the probability of agreement by chance that is more conservative by comparison to calculation of a simple joint probability of agreement percentage.

4.2.4 Theoretical framework

After the thematic analysis stage has produced a reliable coding template, this structure can be used as a theoretical framework for further investigation and verification. This simply means that the primary top-level themes in the coding template (and sub-themes depending on the level of investigation required) become the guiding framework for further data collection, i.e. the topics form the questions that will be asked in the next quantitative line of enquiry, via questionnaire.

4.3 Questionnaire design

4.3.1 General

Template analysis identifies the key characteristics of the context that impact on the psychological dimensions being investigated. These topics can now be taken forward in the initial design of a questionnaire which will be the foundation of the psychometric tool. Thus, in the development of the worker satisfaction measure for automated work systems, the template analysis will have revealed the characteristics of the particular system(s) that were experienced by the participants in relation to their satisfaction, and these that now need to be developed into questions or items (3.3) in the development of the psychometric tool.

4.3.2 Item generation

The psychometric tool will consist of a survey or questionnaire format. Therefore, a set of items should be developed for with care to ensure that responses will provide measures that relate to the key characteristics of the specific situation or context that have been identified. Thus, a pool of preliminary questionnaire items need to be developed that measure the person's level of satisfaction in response to each of the different characteristics identified as relevant to the automated work system(s) context. For example, if speed of automation was a topic that was commonly found in the qualitative data from participant responses then this signifies that speed is a key characteristics that is likely to affect individuals' satisfaction levels, so item(s) are needed for the questionnaire to measure responses related to automation speed.

As the top-level themes identified in the template analysis reflect the most important characteristics of the given context, similarly the sub-themes of the coding template reflect individual dimensions of these characteristics. For example, if the template analysis identifies that speed of automation is a top-level theme for the context, then speed is likely to consist of a number of related dimensions that affect outcomes, e.g. throughput programme, payload, gripper speed, tool centre point, etc. Therefore, at least one item should be created for each identified theme/sub-theme that emerged in the template analysis and taken forward for testing. To avoid response bias, items should not all be phrased in the same positive/negative direction.

4.3.3 Format

When a sufficient pool of items has been collated, they should be assembled into a preliminary questionnaire with a format that is designed so that it is easy to read, understand, and respond to, in accordance with principles of accessibility, i.e. across 'a population with the widest range of user' (ISO 9241-210:2019). Items pertaining to the same theme should be dispersed across the questionnaire and not positioned alongside each other to avoid response bias; for example if two items are intended to measure aspects of automation speed in relation to worker satisfaction it is preferable to position them apart and interspersed by other items. Additionally, if possible, items should test both directions of response for each theme, i.e. items should ask participants to provide scores in response to both negatively and positively phrased statements. If numeric scales are needed the most suitable number of response points should be considered, selected and kept consistent across all items, for example Likert scales may comprise scoring intervals of 5, 7 or more. Any instructions or briefing/debriefing information should also be developed and inserted into the questionnaire.

When the entire set of items and accompanying information has been assembled into a full questionnaire, this should be reviewed for the purposes of checking content and face validity, and ideally also pilot-tested with a small number of participants (unconnected to main studies) to check if it is found sufficiently easy to read, understand and respond to. Any identified errors and omissions can then be addressed by editing the design, and retesting. Researchers should pay particular attention to identifying and correcting any problems with the design of the questionnaire which may impede responses and accuracy, such as clarity of the wording and face validity, or the instructions. This process of trial and revision should be conducted as many times as is necessary to optimise the accessibility of this preliminary questionnaire.

4.4 Experimental study

4.4.1 General

After initial formatting and piloting of the questionnaire, the main experimental study phase of the methodology is necessary to administer, test and analyse results to evaluate the degree to which the items are reliably capturing the data for which they are intended. In contrast to the exploratory study described in section 4.2, the experimental phase does not begin without *a priori* knowledge but with a developed set of items based on empirical data gathered from participants in the exploratory study.

4.4.2 Experimental design

The experimental study should be designed to generate quantitative data for statistical analysis using a structured questionnaire format. Participants should have existing experience or be given new experience of the context that is being examined, so that they can comment on it, before the questionnaire is administered to them. There are potential variations in how/when the questionnaire can be administered but it will typically be provided after the context is experienced for retrospective responses. Thus, for development of the worker satisfaction in automation psychometric tool, the participant would ideally be given experience of the automation context and then immediately administered with the questionnaire so they can recall their perceptions as soon as possible afterwards to minimise any impacts from memory decay.

Sample sizes for experimental research need to be large enough to sustain statistical interrogation. For robust factor analysis and item discrimination in the design of questionnaires, a sample of n = 300 has been recommended (Comrey and Lee, 1992), however this magnitude is unlikely to be attainable in industrial studies. As with the exploratory study, the ideal sample composition and characteristics should reflect those of the target population. However, it may be beneficial to consider a wider set of characteristics within the sample depending on the anticipated future context and potential changing requirements. For example, with more mobile and ageing workforces it may be considered that a sample should include a more diverse mix of ages and abilities than current workforces comprise. Sample sizes for experimental studies should be determined by the particular demands of the analysis, e.g. number of variables, type of variables, etc., but should be sufficient for statistical analyses.

Research ethics protocols should be incorporated into the research study design so that they are followed at all times in accordance with appropriate codes of practice for human participants, as set out by leading psychological associations as described in 4.2.2. Risk assessments should also be carried out to ensure safety.

The design of the experimental study should be tested to check that it is safe, viable and will be successful in achieving the data being sought. The design can be refined and retested until it is finalised.

4.4.3 Quantitative data collection

To develop a psychometric test (3.6) that will be reliable and valid across different applications it is advisable to a conduct a minimum of two initial experimental studies. This will enable collection of quantitative data for preliminary statistical testing of reliability and validity of items in the scale. The studies may differ in a number of ways, for example by sample, task, location, application, etc. Thus, for the development of a questionnaire that will measure satisfaction in relation to automated industrial systems, studies could be conducted with the same system but with different participant samples (e.g. different shift/teams) or different assembly tasks etc., or could be conducted with comparable systems with the same sample or task.

4.4.4 Statistical analysis

Statistical analysis is vital for testing the questionnaire's viability and effectiveness. It is necessary to ensure that each item measures the dimension/sub-dimension that it is intended to measure (and eliminate those that to not), and to examine dimensionality, reliability and validity. For these purposes there are various approaches and statistical procedures that may be used when two or more experimental studies have been conducted with sufficient sample sizes. The following are recommendations of apt procedures and the logical order in which they should be performed. However, investigators must select their methods according to the requirements and design of their particular studies.

The first phase of analysis is inevitably to generate and review simple descriptive statistics and frequencies to evaluate key indications and summaries of the data and get an overall understanding of the dataset. This should then be followed by statistical analysis of differences between the datasets produced by the studies. Standard procedures for analysing differences between the mean scores in two groups that may be used are the t-test which analyses two groups/conditions, and the Analysis of Variance (ANOVA) procedure which tests mean differences across more than two datasets. As ANOVA is a parametric test based on assumptions of interval level data and normal distribution, precursory tests of normality and homogeneity of variance are required.

Preliminary reliability assessment should then be performed using the Cronbach's alpha statistic which reflects internal consistency of a scale. As part of this process it is possible to identify the extent to which the deletion of each item would significantly improve overall reliability of the scale. If there are any items that do not significantly contribute to reliability this means they are not sufficiently related to the scale and, therefore, should be removed.

With unrelated items removed, the dimensionality or structure of the questionnaire scale procedures should be explored with further scrutiny of each individual item and its contribution to the overall scale. This allows further item discrimination to remove 'poor' items. Item discrimination procedures verify the factor structure of individual questionnaire items by quantifying the degree to which each questionnaire item loads onto a specific factor. For example, using the example of satisfaction and speed, item discrimination techniques would quantify how much a specific questionnaire item was measuring satisfaction with speed, or a sub-dimension of speed, in isolation from its relationship with other factors. This process identifies which items are reliable measures for specific factors and should be retained and which are 'poor' items that have low 'factor loadings' that do not contribute to the analytic power of the questionnaire and, therefore, should be eliminated. Reduction of items is advantageous as it reduces the size of a questionnaire without losing its validity and reliability.

There are various techniques that can be employed to identify the common underlying structure of a dataset and evaluate construct validity. Principal Components Analysis (PCA) and Factor Analysis (FA) both generate the linear relationships between items in a dataset to identify the commonality of underlying constructs, and produce similar results. In general, FA is recommended for more theoretical data (e.g. that derived from an unstructured approach when there has been *no existing data* found in the preliminary review of existing knowledge or when the hybrid approach includes data that is not yet known to be reliable, whereas PCA may be more appropriate for an analytical route using *a priori* data that is established as reasonably reliable in measuring factors. Additionally, the dataset may be scrutinised using correlations to explore the two constituent aspects of content validity – convergent validity (similarity) and discriminant validity (difference) – in order to establish the relationship of individual items.

If dimensionality, reliability and validity analyses reveal any problems that need to be addressed refinements to the psychometric scale and further quantitative data collection can proceed until the tool is satisfactory.

Bibliography

[1] ISO 6385:2016, *Ergonomics principles in the design of work systems*

[2] ISO 9241-11:2018, *Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*

[3] ISO 9241-210:2019, *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*

[4] Cronbach and Furby, 1970

[5] Warr, Cook and Wall, 1979

[6] Comrey and Lee, 1992

[7] King, N., "Template Analysis," in Qualitative Methods and Analysis in Organisational Research: A Practical Guide, Symon, G and Cassell, C., (Ed.), Sage London, 1998

[8] Smith and Smith, 2005

[9] Charalambous, G., Fletcher, S., & Webb, P., The development of a scale to evaluate Trust in industrial human-robot collaboration. International Journal of Social Robotics, 8(2), 193-209, 2016

[10] https://www.bps.org.uk/sites/default/files/documents/code_of_human_research_ethics.pdf

[11] http://www.apa.org/ethics/code/index.aspx

[12] http://ethics.efpa.eu/metaand-model-code/meta-code/